

# The Iteration Number of the Weisfeiler-Leman Algorithm

Martin Grohe 

RWTH Aachen University

[grohe@informatik.rwth-aachen.de](mailto:grohe@informatik.rwth-aachen.de)

Moritz Lichter 

TU Darmstadt

[lichter@mathematik.tu-darmstadt.de](mailto:lichter@mathematik.tu-darmstadt.de)

Daniel Neuen 

Simon Fraser University

[dneuen@sfu.ca](mailto:dneuen@sfu.ca)

## Abstract

We prove new upper and lower bounds on the number of iterations the  $k$ -dimensional Weisfeiler-Leman algorithm ( $k$ -WL) requires until stabilization. For  $k \geq 3$ , we show that  $k$ -WL stabilizes after at most  $O(kn^{k-1} \log n)$  iterations (where  $n$  denotes the number of vertices of the input structures), obtaining the first improvement over the trivial upper bound of  $n^k - 1$  and extending a previous upper bound of  $O(n \log n)$  for  $k = 2$  [Lichter et al., LICS 2019].

We complement our upper bounds by constructing  $k$ -ary relational structures on which  $k$ -WL requires at least  $n^{\Omega(k)}$  iterations to stabilize. This improves over a previous lower bound of  $n^{\Omega(k/\log k)}$  [Berkholz, Nordström, LICS 2016].

We also investigate tradeoffs between the dimension and the iteration number of WL, and show that  $d$ -WL, where  $d = \lceil \frac{3(k+1)}{2} \rceil$ , can simulate the  $k$ -WL algorithm using only  $O(k^2 \cdot n^{\lfloor k/2 \rfloor + 1} \log n)$  many iterations, but still requires at least  $n^{\Omega(k)}$  iterations for any  $d$  (that is sufficiently smaller than  $n$ ).

The number of iterations required by  $k$ -WL to distinguish two structures corresponds to the quantifier rank of a sentence distinguishing them in the  $(k+1)$ -variable fragment  $C_{k+1}$  of first-order logic with counting quantifiers. Hence, our results also imply new upper and lower bounds on the quantifier rank required in the logic  $C_{k+1}$ , as well as tradeoffs between variable number and quantifier rank.

## 1 Introduction

The Weisfeiler-Leman (WL) algorithm is a combinatorial algorithm that, given a relational structure  $\mathfrak{A}$  (in most applications, this structure is a graph), iteratively computes an isomorphism-invariant coloring of tuples of vertices of  $\mathfrak{A}$ . The original algorithm introduced by Weisfeiler and Leman [26] is the 2-dimensional version that colors pairs of vertices. Its generalization to arbitrary dimension  $k \geq 1$ , independently introduced by Babai and Mathon as well as Immerman and Lander [11] (see also [1] for a historic note), yields for every natural number  $k$  the  $k$ -dimensional WL algorithm ( $k$ -WL), which iteratively refines a coloring of vertex  $k$ -tuples by aggregating local structural information encoded in the colors. More concretely, the  $k$ -WL algorithm initially colors all  $k$ -tuples of vertices  $\mathbf{v} = (v_1, \dots, v_k)$  of a structure  $\mathfrak{A}$  by the isomorphism type of the underlying induced ordered substructure. Afterwards, in each iteration, the coloring is refined by taking the colors of all tuples into account that can be obtained from  $\mathbf{v}$  by replacing a single entry of the tuple. This process necessarily stabilizes after a finite number of iterations and the resulting coloring can be used to classify  $k$ -tuples of vertices.

The most prominent application of the WL algorithm lies in the context of the graph isomorphism problem. Indeed, since no isomorphism between two structures  $\mathfrak{A}$  and  $\mathfrak{B}$  can map tuples of vertices of different colors to each other, the WL algorithm provides a hierarchy of increasingly powerful heuristics to the graph isomorphism problem. While there is no dimension  $k$  for which

$k$ -WL serves as a complete isomorphism test [4], the algorithm is still surprisingly powerful. For example, Grohe [7] proved that for every non-trivial minor-closed graph class there is some  $k \in \mathbb{N}$  such that  $k$ -WL computes a different coloring on all non-isomorphic graphs, and thus provides a polynomial-time isomorphism test on that class. Moreover, the WL algorithm is also regularly used as a subroutine in isomorphism algorithms (see, e.g., [19, 20, 23]) which includes Babai’s [1] quasipolynomial-time graph isomorphism test that employs the WL algorithm with dimension  $k = O(\log n)$ .

More recently, the WL algorithm has also received significant attention in the machine learning context where it characterizes the expressiveness of graph neural networks [8, 17, 27] and, more generally, the colorings computed by WL are used in classification tasks on graph-structured data sets (see, e.g., [16, 22]).

Since the late 1980s, the WL algorithm has played an important role in descriptive complexity theory. Indeed, it was independently introduced in the context of descriptive complexity by Immerman and Lander [11]. The main reason for this is that  $k$ -WL can be seen as an equivalence test for the logic  $C_{k+1}$ , the  $(k+1)$ -variable fragment of first-order logic with counting quantifiers  $\exists^{\geq n}x$ . Through this connection, the algorithm has turned out to be important for studying the expressiveness of fixed-point logic with counting [4] and, more generally, for the quest for a logic capturing polynomial time [6, 21].

In this work, we study the iteration number of  $k$ -WL, i.e., the number of iterations the algorithm requires until stabilization. Since the number of color classes increases in each iteration, the  $k$ -WL algorithm trivially requires at most  $n^k - 1$  rounds to stabilize. For  $k = 1$ , Kiefer and McKay [12] proved that this trivial bound is optimal by providing several infinite families of graphs  $G$  for which 1-WL requires  $n - 1$  iterations to stabilize (where  $n$  denotes the number of vertices of  $G$ ). In contrast, for  $k = 2$ , Lichter, Ponomarenko and Schweitzer [15] (improving an earlier upper bound by Kiefer and Schweitzer [14]) obtained an upper bound of  $O(n \log n)$  on the iteration number of 2-WL. Beyond that, no improved upper bounds are known for  $k \geq 3$ . As our first main contribution, we obtain non-trivial bounds on the iteration number of  $k$ -WL for all  $k \geq 2$ .

**Theorem 1.1.** *For all  $k \geq 2$ , the  $k$ -dimensional Weisfeiler-Leman algorithm stabilizes after  $O(kn^{k-1} \log n)$  refinement rounds on all relational structures  $\mathfrak{A}$  of arity at most  $k$  where  $n$  denotes the size of the universe.*

For the proof, we extend the algebraic arguments from [15]. Consider a structure  $\mathfrak{A}$  with vertex set  $V$  of size  $n$  and let  $\chi_0, \dots, \chi_\ell: V^k \rightarrow C$  denote the sequence of colorings computed by  $k$ -WL, i.e.,  $\chi_i$  is the coloring computed in the  $i$ -th iteration. For  $k = 2$ , Lichter et al. [15] associate with each coloring  $\chi_i$  a matrix algebra as follows. For each color  $c$  in the image of  $\chi_i$ , let  $M_{i,c}$  denote the  $V \times V$  indicator matrix that sets  $M_{i,c}(v_1, v_2) := 1$  if  $\chi_i(v_1, v_2) = c$ , and  $M_{i,c}(v_1, v_2) := 0$  otherwise. The matrices  $M_{i,c}$ , where  $c$  ranges over all colors in the image of  $\chi_i$ , generate a matrix algebra  $\mathbb{A}^{(i)}$  of  $V \times V$  matrices over the complex numbers using standard matrix multiplication. Using representation-theoretic arguments, it is possible to bound the length of the sequence of matrix algebras generated this way which eventually leads to the upper bound of  $O(n \log n)$ .

The proof of Theorem 1.1 follows a similar strategy. For each color in the image of  $\chi_i$ , we obtain an indicator tensor  $M_{i,c} \in \mathbb{C}^{V^k}$ . Now, the key challenge in generalizing the arguments of [15] is to define a suitable multiplication of those tensors that can be “simulated” by a single round of  $k$ -WL. Given such a multiplication, we then show that the generated algebra  $\mathbb{A}^{(i)}$  is isomorphic to a subalgebra of the  $n^{k-1} \times n^{k-1}$  full matrix algebra (over the complex numbers) which then again allows us to use algebraic arguments to obtain the desired upper bound.

Our arguments actually prove a more general result. Let  $\chi_0, \dots, \chi_\ell: V^k \rightarrow C$  be a sequence of finer and finer colorings (i.e., the partition into color classes of  $\chi_i$  refines the partition into color classes of  $\chi_{i-1}$  for all  $i \in [\ell]$ ) where in each step the coloring is refined at least as much as

by a single iteration of  $k$ -WL. Then the length of the sequence is bounded by  $\ell = O(kn^{k-1} \log n)$ . As a lower bound to our arguments, we show that, in this more general setting, our upper bound is tight up to a factor  $O_k(\log n)$  (the  $O_k(\cdot)$ -notation hides constant factors in  $k$ ). Here, the key insight is that we can find a sequence of finer and finer colorings  $\chi_0, \dots, \chi_\ell: V^k \rightarrow C$  of length  $\Omega_k(n^{k-1})$  that are all stable with respect to  $k$ -WL. As such, it provides a lower bound in the more general setting explained above (but it does not give any lower bounds on the iteration number of  $k$ -WL) and implies that new ideas are likely required to obtain further improvements on the upper bounds of the iteration number of  $k$ -WL (see Section 4 for more details).

Looking for lower bounds on the iteration number of  $k$ -WL, Fürer [5] provided, for every  $k \geq 2$ , a family of graphs on which  $k$ -WL requires at least  $\Omega(n)$  many iterations until stabilization. For  $k$  sufficiently large, this result was strengthened by Berkholz and Nordström [3] who constructed  $k$ -ary relational structures  $\mathfrak{A}$  of size  $n$  on which  $k$ -WL requires at least  $n^{\Omega(k/\log k)}$  many iterations. Answering an open question from [3], our second main contribution is an improved lower bound that gets rid of the  $1/\log k$  factor in the exponent. Actually, we prove the following even stronger result.

**Theorem 1.2.** *There are absolute constants  $k_0 \in \mathbb{N}$  and  $\alpha, \varepsilon > 0$  such that for every  $d \geq k \geq k_0$  and every  $n \geq \alpha d^8 k^6$  there is a pair of  $k$ -ary relational structures  $\mathfrak{A}$  and  $\mathfrak{B}$  of size  $|V(\mathfrak{A})| = |V(\mathfrak{B})| = n$  that are distinguished by  $k$ -WL, but  $d$ -WL does not distinguish  $\mathfrak{A}$  and  $\mathfrak{B}$  after  $n^{\varepsilon k}$  refinement rounds.*

We note that, as in the work of Berkholz and Nordström [3], the structures we need to prove this theorem are  $k$ -ary, that is, have relations of arity  $k$ .

The structures  $\mathfrak{A}$  and  $\mathfrak{B}$  provided by the theorem can be distinguished by  $k$ -WL which trivially requires at most  $n^k - 1$  rounds. The theorem states that, even if we are allowed to increase the dimension of the Weisfeiler-Leman algorithm to  $d$ , the structures can still not be distinguished unless  $d$ -WL runs for at least  $n^{\varepsilon k}$  rounds. This result stands in strong contrast to several existing results for restricted classes of graphs. For example,  $k$ -WL distinguishes between all non-isomorphic pairs of graphs of tree-width at most  $k$  [13], and increasing the dimension to  $4k + 3$  guarantees that  $O(\log n)$  iterations suffices to distinguish between all non-isomorphic pairs of graphs of tree-width at most  $k$  [10]. Similar results are known for planar graphs [9, 25]. The above theorem rules out such results for general relational structures even if we only wish to improve the iteration number to, for example, linear in  $n$ .

By setting  $d = k$ , we obtain the following corollary which shows that the upper bound in Theorem 1.1 is optimal up to a constant factor (that does not depend on  $k$ ) in the exponent.

**Corollary 1.3.** *There are absolute constants  $k_0 \in \mathbb{N}$  and  $\alpha, \varepsilon > 0$  such that for every  $k \geq k_0$  and every  $n \geq \alpha k^{14}$  there is a  $k$ -ary structure  $\mathfrak{A}$  of size  $|V(\mathfrak{A})| = n$  such that the  $k$ -dimensional Weisfeiler-Leman algorithm does not stabilize within  $n^{\varepsilon k}$  refinement rounds on  $\mathfrak{A}$ .*

For the proof of Theorem 1.2, our main technical contribution is to show that there is a  $k_0 \in \mathbb{N}$  such that, for all  $d \geq k_0$ , there are structures  $\mathfrak{A}$  and  $\mathfrak{B}$  of size  $n$  that are distinguished by  $k_0$ -WL, but  $d$ -WL still requires  $\Omega(n/d^2)$  many iterations to distinguish  $\mathfrak{A}$  and  $\mathfrak{B}$ . Afterwards, we obtain Theorem 1.2 by using a known hardness condensation [3] that reduces the size of the structures while roughly preserving the number of iterations required to distinguish them.

Let us point out that Fürer [5] constructed graphs  $G$  and  $H$  which are distinguished by  $k_0$ -WL after  $\Omega(n)$  many rounds. However, as Fürer also shows, his instances are distinguished by  $(3k_0)$ -WL after only  $O(\log n)$  many rounds which means that we cannot use them for our purposes. Berkholz and Nordström [3] provided, for all  $d \geq 2$ , structures  $\mathfrak{A}$  and  $\mathfrak{B}$  of size  $n$  that are distinguished by 2-WL, but  $d$ -WL still requires  $\Omega(n^{1/(1+\log d)})$  many rounds to distinguish them. In combination with the hardness condensation, this leads to the previous lower bound of  $n^{\Omega(k/\log k)}$ .

For the construction of our structures, we introduce the notion of *layered expanders* whose global structure is similar to a  $(k \times n)$ -grid, but that locally (when looking at  $O(k)$  consecutive

columns) behave like an expander graph. We then obtain propositional XOR-formulas from layered expanders which can be transformed into relational structures which satisfy the desired properties.

**Connection to Logics.** As pointed out above,  $k$ -WL is an equivalence test for the logic  $\mathsf{C}_{k+1}$ . That is,  $k$ -WL distinguishes between two structures  $\mathfrak{A}$  and  $\mathfrak{B}$  if and only if there is a sentence  $\varphi \in \mathsf{C}_{k+1}$  such that  $\mathfrak{A} \models \varphi$  and  $\mathfrak{B} \not\models \varphi$ . Additionally, the minimal quantifier rank of such a sentence equals (up to an additive error of at most  $k$ ) the number of iterations  $k$ -WL requires to distinguish between  $\mathfrak{A}$  and  $\mathfrak{B}$ . With this in mind, Theorem 1.1 can be reformulated as follows.

**Corollary 1.4.** *Let  $k \geq 3$ . Let  $\mathfrak{A}$  and  $\mathfrak{B}$  be two relational structures of arity at most  $k$  that can be distinguished by a sentence in  $\mathsf{C}_k$ . Then there is a sentence  $\varphi \in \mathsf{C}_k$  of quantifier rank at most  $q = O(kn^{k-2} \log n)$  such that  $\mathfrak{A} \models \varphi$  and  $\mathfrak{B} \not\models \varphi$ .*

Similarly, we can reformulate Theorem 1.2, but here it turns out that we can obtain an even stronger result since the structures constructed in the theorem can already be distinguished in the logic  $\mathsf{L}_{k+1}$ , the  $(k+1)$ -variable fragment of first-order logic *without* counting quantifiers.

**Theorem 1.5.** *There are absolute constants  $k_0 \in \mathbb{N}$  and  $\alpha, \varepsilon > 0$  such that for every  $d \geq k \geq k_0$  and every  $n \geq \alpha d^8 k^6$  there is a pair of  $k$ -ary structures  $\mathfrak{A}$  and  $\mathfrak{B}$  of size  $|V(\mathfrak{A})| = |V(\mathfrak{B})| = n$  that can be distinguished by a sentence in  $k$ -variable first-order logic  $\mathsf{L}_k$ , but satisfy the same sentences in  $\mathsf{L}_d$  and  $\mathsf{C}_d$  up to quantifier rank  $n^{\varepsilon k}$ .*

Hence, we obtain lower bounds for the quantifier rank not only for the logic  $\mathsf{C}_k$ , but also for the logic  $\mathsf{L}_k$ . We stress that the lower bounds on the quantifier rank remain valid even if we arbitrarily increase the number of variables to any number  $d$  (as long as  $d$  is sufficiently far away from the size of the structures). In other words, even if we are allowed to increase the number of variables, we cannot in general hope for significant improvements on the quantifier rank required to distinguish between two structures.

Having said that, our final result shows that at least some improvements on the upper bound are possible if we are allowed to increase the number of variables by roughly a factor of  $3/2$ .

**Theorem 1.6.** *Let  $k \geq 2$ . Let  $\mathfrak{A}$  and  $\mathfrak{B}$  be two relational structures of arity at most  $k$  such that  $n := |V(\mathfrak{A})| = |V(\mathfrak{B})|$ . Also suppose there is a sentence  $\varphi \in \mathsf{C}_{k+1}$  such that  $\mathfrak{A} \models \varphi$  and  $\mathfrak{B} \not\models \varphi$ . Let  $d := \lceil \frac{3(k+1)}{2} \rceil$ . Then there is a sentence  $\psi \in \mathsf{C}_d^{(q)}$  of quantifier rank  $q = O(k^2 \cdot n^{\lfloor k/2 \rfloor + 1} \log n)$  such that  $\mathfrak{A} \models \psi$  and  $\mathfrak{B} \not\models \psi$ .*

**Structure of the Paper.** After introducing the necessary preliminaries in the next section, we prove Theorem 1.1 in Section 3. Afterwards, we prove limitations of our approach to obtain improved upper bounds on the iteration number in Section 4. In Section 5, we obtain the lower bounds on the iteration number of WL and prove Theorems 1.2 and 1.5. Finally, Theorem 1.6 is proved in Section 6.

## 2 Preliminaries

We use  $\mathbb{N} = \{1, 2, 3, \dots\}$  to denote the positive integers. For  $n \in \mathbb{N}$  we write  $[n] := \{1, \dots, n\}$  and  $[0, n] := \{0, \dots, n\}$ .

**Graphs.** We use standard graph notation. A *graph* is a pair  $G = (V(G), E(G))$  with finite vertex set  $V(G)$  and edge set  $E(G)$ . In this paper, all graphs are simple (i.e., there are no loops or multiedges) and undirected. We write  $vw$  to denote an edge  $\{v, w\} \in E(G)$ . The (*open*) *neighborhood* of a vertex  $v \in V(G)$  is the set  $N_G(v) := \{w \in V(G) \mid vw \in E(G)\}$ . The degree

of a vertex, denoted by  $\deg_G(v)$ , is the size of its neighborhood. For  $X \subseteq V(G)$  we define  $N_G(X) := (\bigcup_{v \in X} N_G(v)) \setminus X$  to denote the neighborhood of  $X$ . If the graph  $G$  is clear from context, we usually omit the index  $G$  and simply write  $N(v)$ ,  $\deg(v)$  and  $N(X)$ . For  $X \subseteq V(G)$  we also write  $G[X]$  to denote the subgraph of  $G$  induced by  $X$ .

**Relational Structures.** In this work, we restrict ourselves to relational vocabularies (signatures)  $\sigma = \{R_1, \dots, R_m\}$  where each  $R_i$  is a relation symbol of a prescribed arity  $k_i \geq 1$ . We say that  $\sigma$  has arity at most  $k$  if  $k_i \leq k$  for all  $i \in [m]$ . A  $\sigma$ -structure is a tuple  $\mathfrak{A} = (V(\mathfrak{A}), R_1^{\mathfrak{A}}, \dots, R_m^{\mathfrak{A}})$  where  $V(\mathfrak{A})$  is a finite universe and  $R_i^{\mathfrak{A}} \subseteq (V(\mathfrak{A}))^{k_i}$  is a relation of arity  $k_i$ . In the remainder of this work, we usually do not explicitly refer to the vocabulary underlying a structure  $\mathfrak{A}$ . With this in mind, we say a structure  $\mathfrak{A} = (V(\mathfrak{A}), R_1^{\mathfrak{A}}, \dots, R_m^{\mathfrak{A}})$  has arity at most  $k$  if the underlying vocabulary has arity at most  $k$ .

For  $X \subseteq V(\mathfrak{A})$  we define  $\mathfrak{A}[X]$  to be the induced substructure of  $\mathfrak{A}$  on  $X$ , i.e.,  $\mathfrak{A}[X]$  is the relational structure with  $V(\mathfrak{A}[X]) = X$  and

$$R_i^{\mathfrak{A}[X]} = R_i^{\mathfrak{A}} \cap X^{k_i}$$

for all  $i \in [m]$ . Let  $\mathfrak{B} = (V(\mathfrak{B}), R_1^{\mathfrak{B}}, \dots, R_m^{\mathfrak{B}})$  be a second structure (over the same vocabulary  $\sigma$ ). An isomorphism from  $\mathfrak{A}$  to  $\mathfrak{B}$  is a bijection  $f: V(\mathfrak{A}) \rightarrow V(\mathfrak{B})$  such that, for all  $i \in [m]$  and all  $v_1, \dots, v_{k_i} \in V(\mathfrak{A})$ , it holds that

$$(v_1, \dots, v_{k_i}) \in R_i^{\mathfrak{A}} \iff (f(v_1), \dots, f(v_{k_i})) \in R_i^{\mathfrak{B}}.$$

The structures  $\mathfrak{A}$  and  $\mathfrak{B}$  are isomorphic if there is an isomorphism from  $\mathfrak{A}$  to  $\mathfrak{B}$ .

**Logics.** Next, we cover bounded-variable fragments of first-order logic (with counting quantifiers). Let  $\sigma = \{R_1, \dots, R_m\}$  be a relational vocabulary and suppose  $R_i$  has arity  $k_i \geq 1$ . We write **FO** to denote standard first-order logic defined via the grammar

$$\varphi ::= x_1 = x_2 \mid R_i(x_1, \dots, x_{k_i}) \mid \varphi \wedge \psi \mid \neg \varphi \mid \exists x_1 \varphi$$

for all  $i \in [m]$  and all variables  $x_j \in \mathcal{V}$  where  $\mathcal{V}$  is an infinite set of variables. We write  $\varphi(x_1, \dots, x_k)$  to indicate that the free variables of  $\varphi$  are among the variables  $\{x_1, \dots, x_k\}$ . For a structure  $\mathfrak{A} = (V(\mathfrak{A}), R_1^{\mathfrak{A}}, \dots, R_m^{\mathfrak{A}})$  and  $\mathbf{v} = (v_1, \dots, v_k) \in (V(\mathfrak{A}))^k$  we write  $\mathfrak{A} \models \varphi(\mathbf{v})$  if  $\mathfrak{A}$  is a model of  $\varphi$  when  $x_i$  is interpreted by  $v_i$ .

We define the quantifier rank of a formula  $\varphi \in \mathbf{FO}$  inductively via

- $\text{qr}(x_1 = x_2) = \text{qr}(R_i(x_1, \dots, x_{k_i})) := 0$  for all  $i \in [m]$  and all variables  $x_j \in \mathcal{V}$ ,
- $\text{qr}(\varphi \wedge \psi) := \max(\text{qr}(\varphi), \text{qr}(\psi))$ ,
- $\text{qr}(\neg \varphi) := \text{qr}(\varphi)$ , and
- $\text{qr}(\exists x \varphi) := \text{qr}(\varphi) + 1$  for all  $x \in \mathcal{V}$ .

We define first-order logic with counting quantifiers **C** to be the extension of **FO** by counting quantifiers of the form  $\exists^{\geq j} x \varphi$ . The formula  $\exists^{\geq j} x \varphi$  is satisfied over a structure  $\mathfrak{A}$  if there are at least  $j$  distinct elements  $v \in V(\mathfrak{A})$  that satisfy  $\varphi$ . We extend the definition of the quantifier rank in the natural way by setting  $\text{qr}(\exists^{\geq j} x \varphi) := \text{qr}(\varphi) + 1$  for all  $x \in \mathcal{V}$ .

For  $k \in \mathbb{N}$  we define  $\mathbf{L}_k$  to be the restriction of **FO** to formulas over at most  $k$  variables, i.e., we restrict ourselves to a set of variables  $\mathcal{V}$  of size exactly  $k$ . Similarly, we define  $\mathbf{C}_k$  to be the restriction of **C** to formulas over at most  $k$  variables.

Moreover, for  $q \geq 0$ , we define  $\mathbf{L}_k^{(q)}$  to the restriction of  $\mathbf{L}_k$  to formulas  $\varphi$  of quantifier rank  $\text{qr}(\varphi) \leq q$ . Similarly, we define  $\mathbf{C}_k^{(q)}$  to the restriction of  $\mathbf{C}_k$  to formulas of quantifier rank at most  $q$ .

**The Weisfeiler-Leman Algorithm.** Next, we describe the  $k$ -WL algorithm. While it is most commonly used as a heuristic to graph isomorphism testing, the algorithm can be applied to any relational structure of arity at most  $k$ .

Let  $\chi_1, \chi_2: V^k \rightarrow C$  be colorings of  $k$ -tuples over a finite set  $V$  where  $C$  is some finite set of colors. The coloring  $\chi_1$  *refines*  $\chi_2$ , denoted  $\chi_1 \preceq \chi_2$ , if  $\chi_1(\mathbf{v}) = \chi_1(\mathbf{w})$  implies  $\chi_2(\mathbf{v}) = \chi_2(\mathbf{w})$  for all  $\mathbf{v}, \mathbf{w} \in V^k$ . Observe that  $\chi_1 \preceq \chi_2$  if and only if the partition into color classes of  $\chi_1$  refines the corresponding partition into color classes of  $\chi_2$ . The colorings  $\chi_1$  and  $\chi_2$  are *equivalent*, denoted  $\chi_1 \equiv \chi_2$ , if  $\chi_1 \preceq \chi_2$  and  $\chi_2 \preceq \chi_1$ . Also,  $\chi_1$  *strictly refines*  $\chi_2$ , denoted  $\chi_1 \prec \chi_2$ , if  $\chi_1 \preceq \chi_2$  and  $\chi_1 \not\equiv \chi_2$ .

Let us fix  $k \geq 2$  and consider a relational structure  $\mathfrak{A} = (V(\mathfrak{A}), R_1^{\mathfrak{A}}, \dots, R_m^{\mathfrak{A}})$  of arity at most  $k$ . Let  $\mathbf{v} = (v_1, \dots, v_k) \in (V(\mathfrak{A}))^k$ . We define the *atomic type* of  $\mathbf{v}$ , denoted by  $\text{atp}_{\mathfrak{A}}(\mathbf{v})$ , to be the isomorphism type of the ordered substructure of  $\mathfrak{A}$  that is induced by  $\{v_1, \dots, v_k\}$ . More concretely, for a second structure  $\mathfrak{B} = (V(\mathfrak{B}), R_1^{\mathfrak{B}}, \dots, R_m^{\mathfrak{B}})$  and a tuple  $\mathbf{w} = (w_1, \dots, w_k) \in (V(\mathfrak{B}))^k$ , it holds that  $\text{atp}_{\mathfrak{A}}(\mathbf{v}) = \text{atp}_{\mathfrak{B}}(\mathbf{w})$  if the mapping  $v_i \mapsto w_i$  is an isomorphism from  $\mathfrak{A}[\{v_1, \dots, v_k\}]$  to  $\mathfrak{B}[\{w_1, \dots, w_k\}]$ .

Next, we describe a single refinement step of  $k$ -WL. Let  $V$  be a finite set and let  $\chi: V^k \rightarrow C$  be a coloring of all  $k$ -tuples over  $V$ . We define the coloring  $\text{step}_k(\chi)$  by setting

$$(\text{step}_k(\chi))(\mathbf{v}) := (\chi(\mathbf{v}), \mathcal{M}_{\chi}(\mathbf{v}))$$

for all  $\mathbf{v} = (v_1, \dots, v_k) \in V^k$  where

$$\mathcal{M}_{\chi}(\mathbf{v}) := \left\{ \left\{ (\chi(\mathbf{v}[w/1]), \dots, \chi(\mathbf{v}[w/k])) \mid w \in V \right\} \right\}$$

and  $\mathbf{v}[w/i] := (v_1, \dots, v_{i-1}, w, v_{i+1}, \dots, v_k)$  is the tuple obtained from  $\mathbf{v}$  by replacing the  $i$ -th entry by  $w$  (and  $\{\dots\}$  denotes a multiset). Observe that  $\text{step}_k(\chi) \preceq \chi$ . We say the coloring  $\chi$  is  *$k$ -stable* if  $\text{step}_k(\chi) \equiv \chi$ .

We define the initial coloring computed by  $k$ -WL on the structure  $\mathfrak{A}$  via  $\chi_k^{(0)}[\mathfrak{A}](\mathbf{v}) := \text{atp}_{\mathfrak{A}}(\mathbf{v})$  for all  $\mathbf{v} \in (V(\mathfrak{A}))^k$ . For  $r \geq 0$  we set

$$\chi_k^{(r+1)}[\mathfrak{A}] := \text{step}_k(\chi_k^{(r)}[\mathfrak{A}]).$$

Since  $\chi_k^{(r+1)}[\mathfrak{A}] \preceq \chi_k^{(r)}[\mathfrak{A}]$  for all  $r \geq 0$ , there is some minimal  $r_{\infty} \leq |V|^k - 1$  such that

$$\chi_k^{(r_{\infty})}[\mathfrak{A}] \equiv \chi_k^{(r_{\infty}+1)}[\mathfrak{A}].$$

We say that  $k$ -WL *stabilizes after*  $r_{\infty}$  *rounds on*  $\mathfrak{A}$  and define  $\chi_k^{(\infty)}[\mathfrak{A}] := \chi_k^{(r_{\infty})}[\mathfrak{A}]$  to be the output coloring of  $k$ -WL. Observe that  $\chi_k^{(\infty)}[\mathfrak{A}]$  is a  $k$ -stable coloring.

Now, let  $\mathfrak{B} = (V(\mathfrak{B}), R_1^{\mathfrak{B}}, \dots, R_m^{\mathfrak{B}})$  be a second structure. Let  $r \geq 0$ . We say that  $k$ -WL *distinguishes*  $\mathfrak{A}$  and  $\mathfrak{B}$  *after*  $r$  *rounds* if there is some color  $c$  such that

$$\left| \left\{ \mathbf{v} \in (V(\mathfrak{A}))^k \mid \chi_k^{(r)}[\mathfrak{A}](\mathbf{v}) = c \right\} \right| \neq \left| \left\{ \mathbf{w} \in (V(\mathfrak{B}))^k \mid \chi_k^{(r)}[\mathfrak{B}](\mathbf{w}) = c \right\} \right|.$$

We also say that  $k$ -WL *distinguishes*  $\mathfrak{A}$  and  $\mathfrak{B}$  if there is some integer  $r \geq 0$  such that  $k$ -WL distinguishes  $\mathfrak{A}$  and  $\mathfrak{B}$  after  $r$  rounds. We write  $\mathfrak{A} \simeq_k \mathfrak{B}$  if  $k$ -WL does not distinguish  $\mathfrak{A}$  and  $\mathfrak{B}$ . Note that, if  $k$ -WL distinguishes  $\mathfrak{A}$  and  $\mathfrak{B}$  and  $k$ -WL stabilizes after  $r_{\infty}$  rounds on  $\mathfrak{A}$ , then  $k$ -WL distinguishes  $\mathfrak{A}$  and  $\mathfrak{B}$  after (at most)  $r_{\infty} + 1$  rounds.

The following connections to bounded-variable fragments of first-order logic with counting quantifiers are well-known. Those connections were first proved in [4, 11] for graphs, but the arguments directly generalize to arbitrary relational structures (see, e.g., [7]).



**Theorem 2.1.** *Let  $k \geq 2$ . Also let  $\mathfrak{A}$  and  $\mathfrak{B}$  be structures of arity at most  $k$  and suppose  $\mathbf{v} \in V(\mathfrak{A})^k$  and  $\mathbf{w} \in V(\mathfrak{B})^k$ . Then, for every  $r \geq 0$ , it holds that  $\chi_k^{(r)}[\mathfrak{A}](\mathbf{v}) \neq \chi_k^{(r)}[\mathfrak{B}](\mathbf{w})$  if and only if there is some  $\varphi(\mathbf{x}) \in \mathbb{C}_{k+1}^{(r)}$  such that  $\mathfrak{A} \models \varphi(\mathbf{v})$  and  $\mathfrak{B} \not\models \varphi(\mathbf{w})$ .*

**Corollary 2.2.** *Let  $k \geq 2$ . Also let  $\mathfrak{A}$  and  $\mathfrak{B}$  be structures of arity at most  $k$ .*

*If there is a sentence  $\varphi \in \mathbb{C}_{k+1}^{(r)}$  such that  $\mathfrak{A} \models \varphi$  and  $\mathfrak{B} \not\models \varphi$ , then the  $k$ -dimensional Weisfeiler-Leman algorithm distinguishes  $\mathfrak{A}$  and  $\mathfrak{B}$  after at most  $r$  refinement rounds.*

*If the  $k$ -dimensional Weisfeiler-Leman algorithm distinguishes  $\mathfrak{A}$  and  $\mathfrak{B}$  after  $r$  refinement rounds, then there is a sentence  $\varphi \in \mathbb{C}_{k+1}^{(r+k)}$  such that  $\mathfrak{A} \models \varphi$  and  $\mathfrak{B} \not\models \varphi$ .*

**Algebras.** Finally, we recall some algebraic tools required in this work. We use  $\mathbb{C}$  to denote the complex numbers.

Recall that a  $\mathbb{C}$ -algebra  $\mathbb{A}$  is a ring which is also a  $\mathbb{C}$ -vector space such that  $a \cdot (\mathbf{v}\mathbf{w}) = (a \cdot \mathbf{v})\mathbf{w} = \mathbf{v}(a \cdot \mathbf{w})$  for all  $a \in \mathbb{C}$  and  $\mathbf{v}, \mathbf{w} \in \mathbb{A}$ . Since we restrict our attention to complex numbers, we simply refer to a  $\mathbb{C}$ -algebra as an algebra. In this work, we are interested in matrix algebras where the algebra consists of  $(d \times d)$ -matrices over the complex numbers with standard matrix multiplication as the ring operation. We write  $M_d(\mathbb{C})$  for the full matrix algebra of all  $(d \times d)$ -matrices over the complex numbers. It is a well-known fact that a matrix algebra  $\mathbb{A} \subseteq M_d(\mathbb{C})$ , which is closed under conjugate transposition, is always *semisimple*. Indeed, if  $M$  is in the Jacobson radical of  $\mathbb{A}$ , then so is  $M^*M$ . But  $M^*M$  is diagonalizable (because it is Hermitian) and nilpotent (because the radical is nilpotent [28, Lemma 1.6.6]) and hence,  $M^*M = 0$  and so  $M = 0$ . Then the radical itself is 0, which is one characterization of semisimplicity.

Hence, we can use the following result to bound the length of sequences of strict subalgebras of  $M_d(\mathbb{C})$  that are closed under conjugate transposition.

**Theorem 2.3** ([15, Theorem 5]). *Let  $\mathbb{A}^{(1)} \subset \dots \subset \mathbb{A}^{(\ell)} \subseteq M_d(\mathbb{C})$  be a sequence of semisimple strict subalgebras. Then  $\ell \leq 2d$ .*

A  $*$ -algebra is an algebra with an additional operation  $*$  such that  $(\mathbf{v} + \mathbf{w})^* = \mathbf{v}^* + \mathbf{w}^*$ ,  $(\mathbf{v}\mathbf{w})^* = \mathbf{w}^*\mathbf{v}^*$ ,  $\mathbf{1}^* = \mathbf{1}$  and  $(\mathbf{v}^*)^* = \mathbf{v}$  for all  $\mathbf{v}, \mathbf{w} \in \mathbb{A}$  (where  $\mathbf{1}$  denotes the unit element). Note that  $M_d(\mathbb{C})$  forms a  $*$ -algebra using conjugate transposition.

### 3 Upper Bounds

In this section, we prove Theorem 1.1. Actually, we prove a more general result on the maximal iteration number of any refinement method that is at least as strong as  $k$ -WL.

For the remainder of this section, let us fix some integer  $k \geq 2$ . Let  $V$  be a finite set and let  $\mathcal{P}$  be a partition of  $V^k$ . For two tuples  $\mathbf{v}, \mathbf{w} \in V^k$ , we write  $\mathbf{v} \sim_{\mathcal{P}} \mathbf{w}$  if there is some  $P \in \mathcal{P}$  such that  $\mathbf{v}, \mathbf{w} \in P$  (i.e.,  $\sim_{\mathcal{P}}$  is the equivalence relation with equivalence classes from  $\mathcal{P}$ ).

We say  $\mathcal{P}$  is *compatible with equality* if for all  $P \in \mathcal{P}$ , all tuples  $(v_1, \dots, v_k), (w_1, \dots, w_k) \in P$ , and all  $i, j \in [k]$  it holds that

$$v_i = v_j \iff w_i = w_j.$$

Moreover, the partition  $\mathcal{P}$  is *shufflable* if for every function  $\pi: [k] \rightarrow [k]$  and every pair of tuples  $(v_1, \dots, v_k), (w_1, \dots, w_k) \in V^k$  it holds that

$$(v_1, \dots, v_k) \sim_{\mathcal{P}} (w_1, \dots, w_k) \implies (v_{\pi(1)}, \dots, v_{\pi(k)}) \sim_{\mathcal{P}} (w_{\pi(1)}, \dots, w_{\pi(k)}). \quad (1)$$

**Observation 3.1.** *Let  $\mathcal{P}$  be a shufflable partition of  $V^k$ . Then*

$$P^\pi := \{(v_{\pi(1)}, \dots, v_{\pi(k)}) \mid (v_1, \dots, v_k) \in P\} \in \mathcal{P}$$

*for every bijection  $\pi: [k] \rightarrow [k]$  and every  $P \in \mathcal{P}$ .*

*Proof.* Let  $Q \in \mathcal{P}$  such that  $Q \cap P^\pi \neq \emptyset$ . This means there is some  $(v_1, \dots, v_k) \in P$  such that  $(v_{\pi(1)}, \dots, v_{\pi(k)}) \in Q$ . Let  $(w_1, \dots, w_k) \in P$  be another tuple. Then  $(v_1, \dots, v_k) \sim_{\mathcal{P}} (w_1, \dots, w_k)$  and thus,  $(v_{\pi(1)}, \dots, v_{\pi(k)}) \sim_{\mathcal{P}} (w_{\pi(1)}, \dots, w_{\pi(k)})$  by Equation (1). Since  $(v_{\pi(1)}, \dots, v_{\pi(k)}) \in Q$ , it follows that  $(w_{\pi(1)}, \dots, w_{\pi(k)}) \in Q$ . So  $P^\pi \subseteq Q$ .

By the same argument,  $Q^{\pi^{-1}} \subseteq P$  which implies that  $Q \subseteq P^\pi$ . Together, this means that  $P^\pi = Q \in \mathcal{P}$ .  $\square$

We say a coloring  $\chi: V^k \rightarrow C$  of  $k$ -tuples is *compatible with equality* if the corresponding partition  $\mathcal{P}$  into color classes is compatible with equality. Similarly,  $\chi$  is *shufflable* if  $\mathcal{P}$  is shufflable.

Recall that  $\text{step}_k(\chi)$  denotes the coloring obtained from  $\chi$  after applying a single refinement round of  $k$ -WL.

**Theorem 3.2.** *Let  $V$  be a finite set of size  $n := |V|$ . Also let  $\chi_0, \dots, \chi_\ell: V^k \rightarrow C$  be a sequence of colorings such that*

- (I)  $\chi_t$  is shufflable and compatible with equality for all  $t \in [0, \ell]$ ,
- (II)  $\text{step}_k(\chi_{t-1}) \succeq \chi_t$  for all  $t \in [\ell]$ , and
- (III)  $\chi_{t-1} \succ \chi_t$  for all  $t \in [\ell]$ .

Then  $\ell \leq 2n^{k-1}(\lceil k \log n \rceil + 1) = O(kn^{k-1} \log n)$ .

Note that Theorem 1.1 immediately follows from Theorem 3.2 by observing that all colorings  $\chi_k^{(i)}[\mathcal{A}]$  obtained from the refinement process of  $k$ -WL are shufflable and compatible with equality.

The proof of Theorem 3.2 relies on algebraic tools. Let  $V$  be a finite set of size  $n := |V|$ . We define a multiplication on the space  $\mathbb{C}^{V^k}$  by

$$(\mathbf{a} \cdot \mathbf{b})(v_1, \dots, v_k) := \sum_{v \in V} \mathbf{a}(v_1, \dots, v_{k-1}, v) \mathbf{b}(v_1, \dots, v_{k-2}, v, v_k) \quad (2)$$

for all  $\mathbf{a}, \mathbf{b} \in \mathbb{C}^{V^k}$ . Note that this multiplication is associative and has a unit  $\mathbf{1}$ , defined by

$$\mathbf{1}(v_1, \dots, v_k) := \begin{cases} 1 & \text{if } v_{k-1} = v_k, \\ 0 & \text{otherwise.} \end{cases}$$

Furthermore, the multiplication is compatible with the vector space structure. Hence, it defines an algebra which we denote by  $\mathbb{A}$ .

With every  $\mathbf{a} \in \mathbb{C}^{V^k}$  we associate a matrix  $M_{\mathbf{a}} \in \mathbb{C}^{V^{k-1} \times V^{k-1}}$  with entries

$$M_{\mathbf{a}}((v_1, \dots, v_{k-1}), (w_1, \dots, w_{k-1})) := \begin{cases} \mathbf{a}(v_1, \dots, v_{k-2}, v_{k-1}, w_{k-1}) & \text{if } v_i = w_i \text{ for all } i \in [k-2], \\ 0 & \text{otherwise.} \end{cases}$$

It is easy to see that the mapping  $\mathbf{a} \mapsto M_{\mathbf{a}}$  is injective and linear. Moreover, it is compatible with multiplication:

$$\begin{aligned} & M_{\mathbf{a}} \cdot M_{\mathbf{b}}((v_1, \dots, v_{k-1}), (w_1, \dots, w_{k-1})) \\ &= \sum_{u_1, \dots, u_{k-1} \in V} M_{\mathbf{a}}((v_1, \dots, v_{k-1}), (u_1, \dots, u_{k-1})) M_{\mathbf{b}}((u_1, \dots, u_{k-1}), (w_1, \dots, w_{k-1})) \\ &= \begin{cases} \sum_{u \in V} \mathbf{a}(v_1, \dots, v_{k-2}, v_{k-1}, u) \mathbf{b}(u_1, \dots, u_{k-2}, u, w_{k-1}) & \text{if } v_i = w_i \text{ for all } i \in [k-2], \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$



$$\begin{aligned}
&= \begin{cases} (\mathbf{a} \cdot \mathbf{b})(v_1, \dots, v_{k-1}, w_{k-1}) & \text{if } v_i = w_i \text{ for all } i \in [k-2], \\ 0 & \text{otherwise} \end{cases} \\
&= M_{\mathbf{a} \cdot \mathbf{b}}((v_1, \dots, v_{k-1}), (w_1, \dots, w_{k-1})).
\end{aligned}$$

And finally,  $M_1$  is the identity matrix. Thus,  $\mathbb{A}$  is isomorphic to a subalgebra of the  $n^{k-1} \times n^{k-1}$ -dimensional matrix algebra  $\mathbb{C}^{V^{k-1} \times V^{k-1}}$ .

For every  $\mathbf{a} \in \mathbb{A}$  we define  $\mathbf{a}^* \in \mathbb{A}$  by

$$\mathbf{a}^*(v_1, \dots, v_k) := \overline{\mathbf{a}(v_1, \dots, v_{k-2}, v_k, v_{k-1})}$$

(here,  $\bar{c}$  denotes the complex conjugate of a number  $c \in \mathbb{C}$ , i.e., if  $c = a + bi$  then  $\bar{c} = a - bi$ ). Then  $M_{\mathbf{a}^*} = (M_{\mathbf{a}})^*$  (the conjugate transpose). Thus,  $*$  is an involution on  $\mathbb{A}$  compatible with the algebra structure, which turns  $\mathbb{A}$  into a  $*$ -algebra.

Since  $\mathbb{A}$  is isomorphic to a subalgebra of  $M_{n^{k-1}}(\mathbb{C})$  which is closed under conjugate transposition, we conclude that  $\mathbb{A}$  is semisimple. Moreover, Theorem 2.3 implies the following corollary.

**Corollary 3.3.** *Let  $\mathbb{A}^{(1)} \subset \dots \subset \mathbb{A}^{(\ell)} \subseteq \mathbb{A}$  be a sequence of semisimple strict subalgebras of  $\mathbb{A}$ . Then  $\ell \leq 2n^{k-1}$ .*

We wish to use the last corollary to obtain an upper bound on the length of the coloring sequence  $\chi_0, \dots, \chi_\ell$  in Theorem 3.2. Towards this end, we associate with every coloring  $\chi_t$  (or the corresponding partition into color classes) a subalgebra of  $\mathbb{A}$  as follows.

For every subset  $A \subseteq \mathbb{C}^{V^k}$ , we let  $\text{span}(A)$  be the linear subspace of  $\mathbb{C}^{V^k}$  generated by  $A$ , and we let  $\langle A \rangle$  be the closure of  $\text{span}(A)$  under multiplication. If  $\mathbf{1} \in \langle A \rangle$ , then  $\langle A \rangle$  is a subalgebra of  $\mathbb{A}$ . As indicated above, we are interested in subalgebras of  $\mathbb{A}$  generated by partitions of the set  $V^k$  in the way explained next.

For every subset  $P \subseteq V^k$ , we define

$$\mathbf{c}_P(\mathbf{v}) := \begin{cases} 1 & \text{if } \mathbf{v} \in P, \\ 0 & \text{otherwise} \end{cases}$$

to be the characteristic vector of  $P$ . For a partition  $\mathcal{P}$  of  $V^k$ , we let  $C_{\mathcal{P}} := \{\mathbf{c}_P \mid P \in \mathcal{P}\}$  and  $\mathbb{A}_{\mathcal{P}} := \langle C_{\mathcal{P}} \rangle$ . If  $\mathbf{1} \in \mathbb{A}_{\mathcal{P}}$ , then  $\mathbb{A}_{\mathcal{P}}$  is a subalgebra of  $\mathbb{A}$ .

**Lemma 3.4.** *Let  $\mathcal{P}$  and  $\mathcal{Q}$  be partitions of  $V^k$  such that  $\mathcal{Q}$  strictly refines  $\mathcal{P}$ . Then  $\text{span}(C_{\mathcal{P}}) \subset \text{span}(C_{\mathcal{Q}})$  and  $\mathbb{A}_{\mathcal{P}} \subseteq \mathbb{A}_{\mathcal{Q}}$ .*

*Proof.* If  $P \in \mathcal{P}$  is the disjoint union of  $Q_1, \dots, Q_m \in \mathcal{Q}$ , then  $\mathbf{c}_P = \sum_{i=1}^m \mathbf{c}_{Q_i}$ . Thus,  $C_{\mathcal{P}} \subseteq \text{span}(C_{\mathcal{Q}})$  and therefore  $\text{span}(C_{\mathcal{P}}) \subseteq \text{span}(C_{\mathcal{Q}})$ . Moreover, there are  $P \in \mathcal{P}, Q \in \mathcal{Q}$  such that  $Q \subset P$ . Then  $\mathbf{c}_Q \notin \text{span}(C_{\mathcal{P}})$ , because all  $\mathbf{a} \in \text{span}(C_{\mathcal{P}})$  are constant on  $P$ . Hence the inclusion is strict.

The second assertion  $\mathbb{A}_{\mathcal{P}} \subseteq \mathbb{A}_{\mathcal{Q}}$  follows immediately from the definitions of  $\mathbb{A}_{\mathcal{P}}$  and  $\mathbb{A}_{\mathcal{Q}}$ .  $\square$

**Observation 3.5.** *Let  $\mathcal{P}$  be a partition of  $V^k$ .*

- (1) *If  $\mathcal{P}$  is compatible with equality, then  $\mathbf{1} \in \mathbb{A}_{\mathcal{P}}$  and hence  $\mathbb{A}_{\mathcal{P}}$  is a subalgebra of  $\mathbb{A}$ .*
- (2) *If  $\mathcal{P}$  is shufflable, then  $\mathbb{A}_{\mathcal{P}}$  is closed under  $*$ .*

*Proof.* Suppose that  $\mathcal{P}$  is compatible with equality. Then there is some  $Q \subseteq \mathcal{P}$  such that  $\mathbf{1} = \sum_{P \in Q} \mathbf{c}_P$ . Hence,  $\mathbf{1} \in \mathbb{A}_{\mathcal{P}}$ .

Next, suppose that  $\mathcal{P}$  is shufflable. Consider the bijection  $\pi: [k] \rightarrow [k]$  for which  $\pi(i) = i$  for all  $i \in [k-2]$ ,  $\pi(k-1) = k$  and  $\pi(k) = k-1$ . Then  $\mathbf{c}_P^* = \mathbf{c}_{P\pi}$  for every  $P \in \mathcal{P}$ . Using Observation 3.1, it follows that  $\mathbf{c}_P^* \in C_{\mathcal{P}}$  which implies that  $\mathbb{A}_{\mathcal{P}}$  is closed under  $*$ .  $\square$

**Corollary 3.6.** *Let  $\mathcal{P}$  be a partition of  $V^k$  that is shufflable and compatible with equality. Then  $\mathbb{A}_{\mathcal{P}}$  is a  $*$ -subalgebra of  $\mathbb{A}$ . In particular,  $\mathbb{A}_{\mathcal{P}}$  is semisimple.*

Recall that our goal is to bound the length of the color sequence  $\chi_0, \dots, \chi_\ell$  in Theorem 3.2. We associate a  $*$ -subalgebra  $\mathbb{A}^{(t)}$  of  $\mathbb{A}$  with every coloring  $\chi_t$  by considering the corresponding partition into color classes. The last corollary implies that  $\mathbb{A}^{(t)}$  is semisimple for every  $t \in [0, \ell]$ . So, to be able to apply Corollary 3.3 to bound the length of the sequence of subalgebras, it remains to argue that inclusions between successive subalgebras are strict. Actually, this is not true in general, but we can prove that only a small number of successive algebras can be equal.

We say that  $\mathbf{a} \in \mathbb{C}^{V^k}$  *distinguishes*  $\mathbf{v}, \mathbf{w} \in V^k$  if  $\mathbf{a}(\mathbf{v}) \neq \mathbf{a}(\mathbf{w})$ , and we say that  $A \subseteq \mathbb{C}^{V^k}$  distinguishes  $\mathbf{v}, \mathbf{w}$  if some  $\mathbf{a} \in A$  distinguishes them.

**Lemma 3.7.** *Let  $A \subseteq \mathbb{C}^{V^k}$  and  $\mathbf{v}, \mathbf{w} \in V^k$  such that  $\langle A \rangle$  distinguishes  $\mathbf{v}, \mathbf{w}$ . Then there are an  $s \leq n^k$  and  $\mathbf{a}_1, \dots, \mathbf{a}_s \in A$  such that  $\mathbf{a}_1 \cdots \mathbf{a}_s$  distinguishes  $\mathbf{v}, \mathbf{w}$ .*

*Proof.* As a linear subspace of  $\mathbb{C}^{V^k}$ , the space  $\langle A \rangle$  consists of finite linear combinations of “monomials”  $\mathbf{a}_1 \cdots \mathbf{a}_s$  for  $\mathbf{a}_i \in A$ . Since the dimension of the space is at most  $n^k$ , we only need to consider such monomials for  $s \leq n^k$ . Hence  $\mathbf{v}, \mathbf{w}$  are distinguished by a linear combination

$$\sum_{i=1}^m \lambda_i \mathbf{a}_{i1} \cdots \mathbf{a}_{is_i}$$

with  $\lambda_i \in \mathbb{C}$ ,  $\mathbf{a}_{ij} \in A$ , and  $s_i \leq n^k$ . This immediately implies that  $\mathbf{v}, \mathbf{w}$  are distinguished by  $\mathbf{a}_{i1} \cdots \mathbf{a}_{is_i}$  for some  $i \in [m]$ .  $\square$

With every partition  $\mathcal{P} = \{P_1, \dots, P_m\}$  we associate a relational structure  $(V, R_1^{\mathcal{P}}, \dots, R_m^{\mathcal{P}})$  whose vocabulary consists of  $k$ -ary relation symbols  $R_i$  interpreted by  $R_i^{\mathcal{P}} = P_i$  (to uniquely define the associated structure, we fix an arbitrary order on the blocks  $P_1, \dots, P_m$ ). Slightly abusing notation, we denote this structure by  $\mathcal{P}$  as well. We say that a formula  $\varphi(\mathbf{x})$  *distinguishes*  $\mathbf{v}, \mathbf{w} \in V^k$  over  $\mathcal{P}$  if

$$\mathcal{P} \models \varphi(\mathbf{v}) \iff \mathcal{P} \not\models \varphi(\mathbf{w}).$$

Recall that  $\mathbb{C}_{k+1}^{(q)}$  denotes the fragment of first-order logic with counting consisting of all formulas of quantifier rank at most  $q$  with at most  $k+1$  variables.

**Lemma 3.8.** *Let  $\mathcal{P}$  be a partition of  $V^k$  and let  $\mathbf{v}, \mathbf{w} \in V^k$  such that  $\mathbb{A}_{\mathcal{P}}$  distinguishes  $\mathbf{v}$  and  $\mathbf{w}$ . Then there is a formula  $\varphi(\mathbf{x}) \in \mathbb{C}_{k+1}^{(q)}$  of quantifier rank  $q \leq \lceil k \log n \rceil$  that distinguishes  $\mathbf{v}, \mathbf{w}$  over  $\mathcal{P}$ .*

*Proof.* Suppose that  $\mathcal{P} = \{P_1, \dots, P_m\}$ , and let  $\mathbf{c}_i := \mathbf{c}_{P_i}$ . Then  $\mathbb{A}_{\mathcal{P}} = \langle \{\mathbf{c}_1, \dots, \mathbf{c}_m\} \rangle$ . Thus, by Lemma 3.7, there is an  $s \leq n^k$  and  $i_1, \dots, i_s \in [m]$  such that  $\mathbf{c}_{i_1} \cdots \mathbf{c}_{i_s}$  distinguishes  $\mathbf{v}, \mathbf{w}$ .

By induction on  $s \geq 1$ , we prove that if  $\mathbf{c}_{i_1} \cdots \mathbf{c}_{i_s}$  distinguishes  $\mathbf{v}, \mathbf{w}$ , then there is a formula  $\varphi(\mathbf{x}) \in \mathbb{C}_{k+1}^{(\lceil \log s \rceil)}$  that distinguishes  $\mathbf{v}, \mathbf{w}$ . The assertion of the lemma follows.

For the base step  $s = 1$ , note that if  $\mathbf{c}_i$  distinguishes  $\mathbf{v}, \mathbf{w}$ , then the atomic formula  $R_i(\mathbf{x})$  distinguishes  $\mathbf{v}, \mathbf{w}$ .

For the inductive step, let  $s \geq 2$ . Suppose that  $\mathbf{b} = \mathbf{c}_{i_1} \cdots \mathbf{c}_{i_s}$  distinguishes  $\mathbf{v}, \mathbf{w}$ . Let  $r := \lceil s/2 \rceil$  and note that  $r \leq 2^{\lceil \log s \rceil - 1}$  and therefore

$$\lceil \log r \rceil \leq \lceil \log s \rceil - 1.$$

Let  $\mathbf{b}_1 := \mathbf{c}_{i_1} \cdots \mathbf{c}_{i_r}$  and  $\mathbf{b}_2 := \mathbf{c}_{i_{r+1}} \cdots \mathbf{c}_{i_s}$ . Then  $\mathbf{b} = \mathbf{b}_1 \cdot \mathbf{b}_2$ . Suppose that  $\mathbf{v} = (v_1, \dots, v_k)$  and  $\mathbf{w} = (w_1, \dots, w_k)$ . We have

$$\mathbf{b}(\mathbf{v}) = \sum_{u \in V} \mathbf{b}_1(v_1, \dots, v_{k-1}, u) \cdot \mathbf{b}_2(v_1, \dots, v_{k-2}, u, v_k)$$

$$\neq \mathbf{b}(\mathbf{w}) = \sum_{u \in V} \mathbf{b}_1(w_1, \dots, w_{k-1}, u) \cdot \mathbf{b}_2(w_1, \dots, w_{k-2}, u, w_k).$$

Thus, there are  $b_1, b_2 \in \mathbb{C}$  such that

$$p := \left| \left\{ u \in V \mid \mathbf{b}_1(v_1, \dots, v_{k-1}, u) = b_1 \text{ and } \mathbf{b}_2(v_1, \dots, v_{k-2}, u, v_k) = b_2 \right\} \right| \\ \neq \left| \left\{ u \in V \mid \mathbf{b}_1(w_1, \dots, w_{k-1}, u) = b_1 \text{ and } \mathbf{b}_2(w_1, \dots, w_{k-2}, u, w_k) = b_2 \right\} \right| =: q.$$

It follows from the induction hypothesis that for  $i = 1, 2$  and for all  $\mathbf{v}', \mathbf{w}' \in V^k$  such that  $\mathbf{b}_i$  distinguishes  $\mathbf{v}', \mathbf{w}'$  there is a formula  $\psi_i^{\mathbf{v}', \mathbf{w}'}(\mathbf{x}) \in \mathbb{C}_{k+1}^{(\lceil \log r \rceil)}$  that distinguishes  $\mathbf{v}', \mathbf{w}'$ . Without loss of generality,

$$\mathcal{P} \models \psi_i^{\mathbf{v}', \mathbf{w}'}(\mathbf{v}') \quad \text{and} \quad \mathcal{P} \not\models \psi_i^{\mathbf{v}', \mathbf{w}'}(\mathbf{w}'),$$

otherwise we replace  $\psi_i^{\mathbf{v}', \mathbf{w}'}(\mathbf{x})$  by its negation. Let  $V_i \subseteq V^k$  be the set of all  $\mathbf{v}' \in V^k$  such that  $\mathbf{b}_i(\mathbf{v}') = b_i$  and let

$$\varphi_i(\mathbf{x}) := \bigvee_{\mathbf{v}' \in V_i} \bigwedge_{\mathbf{w}' \in V^k \setminus V_i} \psi_i^{\mathbf{v}', \mathbf{w}'}(\mathbf{x}).$$

Then for all  $\mathbf{v}' \in V^k$  we have

$$\mathcal{P} \models \varphi_i(\mathbf{v}') \iff \mathbf{b}_i(\mathbf{v}') = b_i.$$

Without loss of generality we assume that  $p > q$ . Then the formula

$$\varphi(x_1, \dots, x_k) := \exists^{\geq p} x_{k+1} (\varphi_1(x_1, \dots, x_{k-1}, x_{k+1}) \wedge \varphi_2(x_1, \dots, x_{k-2}, x_{k+1}, x_k)) \in \mathbb{C}_{k+1}^{(\lceil \log s \rceil)}$$

distinguishes  $\mathbf{v}, \mathbf{w}$ . □

We are now ready to prove Theorem 3.2.

*Proof of Theorem 3.2.* For every  $t \in [0, \ell]$  let  $\mathcal{P}^{(t)}$  be the partition of  $V^k$  into the color classes of  $\chi_t$ .

*Claim 3.9.* Let  $t, q \geq 0$  such that  $t + q \leq \ell$ . Suppose that there is a formula  $\varphi(\mathbf{x}) \in \mathbb{C}_{k+1}^{(q)}$  that distinguishes  $\mathbf{v}, \mathbf{w} \in V^k$  over  $\mathcal{P}^{(t)}$ . Then  $\mathbf{v}, \mathbf{w}$  belong to different classes of the partition  $\mathcal{P}^{(t+q)}$ .

*Proof.* By Condition (I), the partition  $\mathcal{P}^{(t)}$  is shufflable and compatible with equality. This implies that  $\chi_t \equiv \chi_k^{(0)}[\mathcal{P}^{(t)}]$ . Together with Condition (II), we get that  $\chi_{t+q} \preceq \chi_k^{(q)}[\mathcal{P}^{(t)}]$ .

Also, using Theorem 2.1, we get that  $\chi_k^{(q)}[\mathcal{P}^{(t)}](\mathbf{v}) \neq \chi_k^{(q)}[\mathcal{P}^{(t)}](\mathbf{w})$ . Overall, it follows that  $\mathbf{v}, \mathbf{w}$  belong to different classes of the partition  $\mathcal{P}^{(t+q)}$ . ┘

For every  $t \in [0, \ell]$  we define  $C^{(t)} := C_{\mathcal{P}^{(t)}}$  and  $\mathbb{A}^{(t)} := \mathbb{A}_{\mathcal{P}^{(t)}}$ . Note that  $\mathbb{A}^{(t)}$  is a semisimple  $*$ -subalgebra of  $\mathbb{A}$  by Condition (I) and Corollary 3.6. By Lemma 3.4, we have

$$\mathbb{A}^{(0)} \subseteq \mathbb{A}^{(1)} \subseteq \dots \subseteq \mathbb{A}^{(\ell)} \subseteq \mathbb{A}. \quad (3)$$

*Claim 3.10.* For all  $t \in [0, \ell - \lceil k \log n \rceil]$ ,

$$\mathbb{A}^{(t)} \subseteq \text{span}(C^{(t+\lceil k \log n \rceil)}).$$

*Proof.* Let  $\mathbf{a} \in \mathbb{A}^{(t)}$ . By Lemma 3.8 and Claim 3.9, for all  $\mathbf{v}, \mathbf{w} \in V^k$ , if  $\mathbf{a}(\mathbf{v}) \neq \mathbf{a}(\mathbf{w})$ , that is, if  $\mathbf{a}$  distinguishes  $\mathbf{v}$  and  $\mathbf{w}$ , then  $\mathbf{v}$  and  $\mathbf{w}$  belong to different classes of the partition  $\mathcal{P}^{(t+\lceil k \log n \rceil)}$ . Thus,  $\mathbf{a}$  is constant on each class of the partition  $\mathcal{P}^{(t+\lceil k \log n \rceil)}$ , which immediately implies that  $\mathbf{a}$  can be written as a linear combination of the characteristic vectors  $\mathbf{c}_P$  of the classes  $P \in \mathcal{P}^{(t+\lceil k \log n \rceil)}$ . This is the assertion of the claim. ┘

*Claim 3.11.* For all  $t \in [0, \ell - \lceil k \log n \rceil - 1]$ ,

$$\mathbb{A}^{(t)} \subset \mathbb{A}^{(t + \lceil k \log n \rceil + 1)}.$$

*Proof.* By Claim 3.10, we have  $\mathbb{A}^{(t)} \subseteq \text{span}(C^{(t + \lceil k \log n \rceil)})$ . Moreover, by Condition (III), the partition  $\mathcal{P}^{(t + \lceil k \log n \rceil + 1)}$  strictly refines the partition  $\mathcal{P}^{(t + \lceil k \log n \rceil)}$ . By Lemma 3.4, this implies  $\text{span}(C^{(t + \lceil k \log n \rceil)}) \subset \text{span}(C^{(t + \lceil k \log n \rceil + 1)})$ . As  $\text{span}(C^{(t + \lceil k \log n \rceil + 1)}) \subseteq \mathbb{A}^{(t + \lceil k \log n \rceil + 1)}$ , the assertion of the claim follows.  $\square$

Recall that by Corollary 3.6, the algebras  $\mathbb{A}^{(t)}$  are semisimple. Thus, by Corollary 3.3, at most  $2n^{k-1}$  of the inclusions in (3) are strict. Then Claim 3.11 implies

$$\ell \leq 2n^{k-1}(\lceil k \log n \rceil + 1) = O(kn^{k-1} \log n). \quad \square$$

## 4 Long Sequences of Stable Colorings

Next, we prove an almost matching lower bound for Theorem 3.2, i.e., we prove that there are sequences of colorings  $\chi_0, \dots, \chi_\ell: V^k \rightarrow C$  satisfying Conditions (I) - (III) of Theorem 3.2 of length  $\ell = \Omega(n^{k-1})$ . Actually, we prove a slightly stronger result.

As before, let us fix an integer  $k \geq 2$ . We present a construction for a sequence  $\chi_0 \succ \chi_1 \succ \dots \succ \chi_\ell$  of colorings of  $V^k$  such that  $\chi_t$  is  $k$ -stable (i.e., the coloring is stable with respect to  $k$ -WL) for all  $t \in [0, \ell]$ . More precisely, the main result of this section is the following theorem.

**Theorem 4.1.** *Suppose  $n \geq 2k^2$  and let  $V$  be a set of size  $|V| = 2n$ . Then there is a sequence of colorings  $\chi_0, \dots, \chi_\ell: V^k \rightarrow C$  of length  $\ell \geq \left(\frac{n}{2k}\right)^{k-1}$  such that*

- (I)  $\chi_t$  is shufflable and compatible with equality for all  $t \in [0, \ell]$ ,
- (II)  $\chi_t$  is  $k$ -stable for all  $t \in [0, \ell]$ , and
- (III)  $\chi_{t-1} \succ \chi_t$  for all  $t \in [\ell]$ .

Before diving into the proof, let us first discuss some implications of the theorem.

First of all, Theorem 4.1 implies that the upper bound in Theorem 3.2 is tight up to a factor of  $O_k(\log n)$ . This follows from the simple observation that, if  $\chi_t$  is  $k$ -stable and  $\chi_{t-1} \succ \chi_t$ , then  $\text{step}_k(\chi_{t-1}) \equiv \chi_{t-1} \succeq \chi_t$ , i.e., the sequence of colorings constructed in Theorem 4.1 satisfies the requirements of Theorem 3.2.

On the other hand, since all colorings  $\chi_t$  are already  $k$ -stable, the theorem does not provide any lower bounds on the iteration number of  $k$ -WL. However, Theorem 4.1 still provides some valuable insights in this setting. Indeed, all existing methods to bound the iteration number of  $k$ -WL [14, 15] rely on “parallelization arguments”, i.e., it is argued that at some point in the refinement process many color classes have to be split at the same time. Theorem 4.1 essentially implies that such arguments do not suffice to push the upper bounds on the iteration number beyond  $O(n^{k-1})$  since such “parallelization arguments” typically also work in the extended setting of Theorem 3.2. As a concrete example, Kiefer and Schweitzer [14] prove upper bounds on iteration number of 2-WL by bounding the cost of a certain game related to 2-WL. This game naturally generalizes to  $k$ -WL, but Theorem 4.1 immediately implies that its cost is  $\Omega(n^{k-1})$  and thus, it is not possible to obtain improved upper bounds by analyzing said game. So overall, Theorem 4.1 can be interpreted as saying that, in order to obtain improved upper bounds on the iteration number of  $k$ -WL, we need to rely on arguments that also exploit the possibility of stabilization at an early point, and it is not possible to solely rely on “parallelization arguments”.

Let us now turn to the proof of Theorem 4.1. It relies on the following theorem which provides a large set family with restricted intersections between its members. Let  $U$  be a set of size  $n$ . A  $k$ -uniform set family (over  $U$ ) is a collection  $\mathcal{F}$  of  $k$ -element subsets of  $U$ .

**Theorem 4.2** ([2, Theorem 4.11]). *For every  $n \geq 2k^2$  there exists a  $k$ -uniform set family  $\mathcal{F}$  over a universe  $U$  of  $n$  points such that*

1.  $|E_1 \cap E_2| \leq k - 2$  for all distinct  $E_1, E_2 \in \mathcal{F}$ , and
2.  $|\mathcal{F}| \geq \left(\frac{n}{2k}\right)^{k-1}$ .

Now, let  $U$  be a universe of size  $n \geq k$  and let  $\mathcal{F}$  be a  $k$ -uniform set family over  $U$ . We set

$$V := U \times \{0, 1\}$$

and define a coloring  $\chi_{\mathcal{F}}: V^k \rightarrow C$  as follows. Since the actual names of the colors are not relevant for our purposes, we only define the color classes, i.e., we specify when two tuples receive the same color.

Let  $((u_1, a_1), \dots, (u_k, a_k)), ((u'_1, a'_1), \dots, (u'_k, a'_k)) \in V^k$ . We define  $\chi_{\mathcal{F}}$  in such a way that  $\chi_{\mathcal{F}}((u_1, a_1), \dots, (u_k, a_k)) = \chi_{\mathcal{F}}((u'_1, a'_1), \dots, (u'_k, a'_k))$  if and only if

- (A)  $u_i = u'_i$  for all  $i \in [k]$ ,
- (B)  $(u_i, a_i) = (u_j, a_j) \Leftrightarrow (u'_i, a'_i) = (u'_j, a'_j)$  for all  $i, j \in [k]$ , and
- (C) if  $\{u_1, \dots, u_k\} \in \mathcal{F}$ , then  $\sum_{i \in [k]} a_i \equiv \sum_{i \in [k]} a'_i \pmod{2}$ .

**Lemma 4.3.** *Suppose  $|E_1 \cap E_2| \leq k - 2$  for all distinct  $E_1, E_2 \in \mathcal{F}$ . Then  $\chi_{\mathcal{F}}$  is  $k$ -stable.*

*Proof.* Let  $\mathbf{v} = ((u_1, a_1), \dots, (u_k, a_k)), \mathbf{v}' = ((u'_1, a'_1), \dots, (u'_k, a'_k)) \in V^k$  such that

$$\chi_{\mathcal{F}}(\mathbf{v}) = \chi_{\mathcal{F}}(\mathbf{v}').$$

Observe that  $u_i = u'_i$  for all  $i \in [k]$  by Condition (A). We need to show that the two tuples do not receive distinct colors after a single refinement step of  $k$ -WL, that is, we need to argue that

$$\begin{aligned} & \left\{ \left( \chi_{\mathcal{F}}(\mathbf{v}[(u, a)/1]), \dots, \chi_{\mathcal{F}}(\mathbf{v}[(u, a)/k]) \right) \mid u \in U, a \in \{0, 1\} \right\} \\ &= \left\{ \left( \chi_{\mathcal{F}}(\mathbf{v}'[(u, a)/1]), \dots, \chi_{\mathcal{F}}(\mathbf{v}'[(u, a)/k]) \right) \mid u \in U, a \in \{0, 1\} \right\} \end{aligned}$$

where  $\mathbf{v}[(u, a)/i] = ((u_1, a_1), \dots, (u_{i-1}, a_{i-1}), (u, a), (u_{i+1}, a_{i+1}), \dots, (u_k, a_k))$  is the tuple obtained from  $\mathbf{v}$  by replacing the  $i$ -th entry by  $(u, a)$ . Towards this end, we actually show the stronger statement that

$$\begin{aligned} & \left\{ \left( \chi_{\mathcal{F}}(\mathbf{v}[(u, a)/1]), \dots, \chi_{\mathcal{F}}(\mathbf{v}[(u, a)/k]) \right) \mid a \in \{0, 1\} \right\} \\ &= \left\{ \left( \chi_{\mathcal{F}}(\mathbf{v}'[(u, a)/1]), \dots, \chi_{\mathcal{F}}(\mathbf{v}'[(u, a)/k]) \right) \mid a \in \{0, 1\} \right\} \end{aligned}$$

holds for all  $u \in U$ .

Fix some  $u \in U$ . To see that these two multisets are equal, consider the set

$$V' := \{u_1, \dots, u_k, u\} \times \{0, 1\} \subseteq V$$

and the restriction  $\chi'_{\mathcal{F}}: (V')^k \rightarrow C: \mathbf{v} \mapsto \chi_{\mathcal{F}}(\mathbf{v})$  of  $\chi_{\mathcal{F}}$  to the set  $(V')^k$ . Also, let  $\mathcal{F}' := \{E \in \mathcal{F} \mid E \subseteq V'\}$ . Since  $|E_1 \cap E_2| \leq k - 2$  for all distinct  $E_1, E_2 \in \mathcal{F}$  and  $|V'| \leq k + 1$ , we conclude that  $|\mathcal{F}'| \leq 1$ .

*Claim 4.4.* There is a bijection  $\varphi: V' \rightarrow V'$  such that

- (i)  $\chi'_{\mathcal{F}}(\mathbf{v}) = \chi'_{\mathcal{F}}(\varphi(\mathbf{v}))$  for all  $\mathbf{v} \in (V')^k$ , and
- (ii)  $\varphi(u_i, a_i) = (u'_i, a'_i)$  for all  $i \in [k]$ .

*Proof.* For  $i \in [k]$  we define  $\varphi(u_i, a_i) := (u_i, a'_i)$  and  $\varphi(u_i, 1 - a_i) := (u_i, 1 - a'_i)$ . In particular, Condition (ii) is satisfied since  $u_i = u'_i$  for all  $i \in [k]$ . If there is some  $E' \in \mathcal{F}'$  such that  $u \in E'$ , then we define

$$\varphi(u, a) := \begin{cases} (u, a) & \text{if } \sum_{u_i \in E'} a_i \equiv \sum_{u_i \in E'} a'_i \pmod{2} \\ (u, 1 - a) & \text{otherwise} \end{cases}$$

for both  $a \in \{0, 1\}$ . If no such set  $E' \in \mathcal{F}'$  exists, then we set  $\varphi(u, a) := (u, a)$  for both  $i \in \{0, 1\}$ . It can be easily verified that  $\chi'_{\mathcal{F}}(\mathbf{v}) = \chi'_{\mathcal{F}}(\varphi(\mathbf{v}))$  for all  $\mathbf{v} \in (V')^k$ .  $\square$

Since the multisets above are defined in an isomorphism-invariant manner over the structure induced by  $(V', \chi'_F)$ , we conclude that they have to be equal.  $\square$

*Proof of Theorem 4.1.* Let  $\mathcal{F}$  be the set family obtained from Theorem 4.2 and suppose that  $\mathcal{F} = \{E_1, \dots, E_\ell\}$ . Observe that  $\ell \geq \left(\frac{n}{2k}\right)^{k-1}$  as desired. For  $t \in [0, \ell]$  we define  $\mathcal{F}_t := \{E_1, \dots, E_t\} \subseteq \mathcal{F}$  and  $\chi_t := \chi_{\mathcal{F}_t}$ . Then  $\chi_t$  is  $k$ -stable by Lemma 4.3. Also,  $\mathcal{F}_{t-1} \subset \mathcal{F}_t$  which implies that  $\chi_{t-1} \succ \chi_t$  by definition of the coloring  $\chi_t$ . Finally, it is easy to verify that all colorings are shufflable and compatible with equality.  $\square$

## 5 Lower Bounds on the Iteration Number of WL

In this section, we obtain improved lower bounds on the iteration number of the Weisfeiler-Leman algorithm. More precisely, we prove Theorem 1.2. Our proof strategy is similar to the one employed by Berkholz and Nordström in [3]. First, for every sufficiently large  $\ell_{\text{hi}} \geq \ell_{\text{lo}}$ , we construct pairs of structures that can be distinguished by  $\ell_{\text{lo}}$ -WL, but  $\ell_{\text{hi}}$ -WL still requires a linear number of iterations to distinguish them. Afterwards, we apply a *hardness compression* that reduces the number of vertices in the obtained structures while preserving the iteration number of the Weisfeiler-Leman algorithm. Actually, for the second step, we can rely on the same tools that are already used by Berkholz and Nordström in [3].

### 5.1 Overview

The hard instances we construct are based on propositional XOR-formulas that can also be viewed as systems of linear equations over the 2-element field  $\mathbb{F}_2$ .

Let  $V$  be a finite set which we interpret as a set of variables that take values in  $\{0, 1\}$ . An *XOR-constraint* (over  $V$ ) is a pair  $(C, a)$  where  $C \subseteq V$  and  $a \in \{0, 1\}$ . The reader is encouraged to think of such a constraint as the equation  $x_1 + \dots + x_k \equiv a \pmod{2}$  where  $C = \{x_1, \dots, x_k\}$  is the set of those variables that appear on the left side of the equation. We explicitly allow  $C$  to be empty;  $(\emptyset, 0)$  is always satisfied and  $(\emptyset, 1)$  is unsatisfiable. Let  $\mathcal{C}$  be a set of XOR-constraints. We define the *arity* of  $\mathcal{C}$  to be the maximum cardinality of  $C$  for any pair  $(C, a) \in \mathcal{C}$ .

We can translate a set of XOR-constraints into a pair of relational structures as follows. Let  $\mathcal{C}$  be a set of XOR-constraints over a set  $V$ . Also suppose that  $V = \{x_1, \dots, x_n\}$ . We define  $\mathfrak{A} = \mathfrak{A}(\mathcal{C})$  and  $\mathfrak{B} = \mathfrak{B}(\mathcal{C})$  as follows. We set  $V(\mathfrak{A}) = V(\mathfrak{B}) := V \times \{0, 1\}$ , i.e., each element of the structures  $\mathfrak{A}$  and  $\mathfrak{B}$  corresponds to an assignment of a single variable. For each  $i \in [n]$ , we add a unary relation  $X_i$  and set  $X_i^{\mathfrak{A}} = X_i^{\mathfrak{B}} := \{(x_i, 0), (x_i, 1)\}$ . Finally, for every constraint  $(C, a) \in \mathcal{C}$  with  $C = \{x_{i_1}, \dots, x_{i_k}\}$  we introduce a  $k$ -ary relation  $R_{C,a}$  and define

$$R_{C,a}^{\mathfrak{A}} := \left\{ \left( (x_{i_1}, b_1), \dots, (x_{i_k}, b_k) \right) \mid b_1, \dots, b_k \in \{0, 1\}, \sum_{j=1}^k b_j \equiv 0 \pmod{2} \right\}$$

and

$$R_{C,a}^{\mathfrak{B}} := \left\{ \left( (x_{i_1}, b_1), \dots, (x_{i_k}, b_k) \right) \mid b_1, \dots, b_k \in \{0, 1\}, \sum_{j=1}^k b_j \equiv a \pmod{2} \right\}$$



Instead of analysing the Weisfeiler-Leman algorithm directly on  $\mathfrak{A}(\mathcal{C})$  and  $\mathfrak{B}(\mathcal{C})$ , it turns out to be more convenient to consider the following game that is directly played on  $\mathcal{C}$  and is known to capture the same information as applying the Weisfeiler-Leman algorithm to the associated structures.

Let  $\mathcal{C}$  be a set of XOR-constraints over a set  $V$ . Let  $k \in \mathbb{N}$  such that  $\mathcal{C}$  has arity at most  $k$ . A partial assignment  $\beta: X \rightarrow \{0, 1\}$  with  $X \subseteq V$  *violates* an XOR-constraint  $(C, a) \in \mathcal{C}$  if  $C \subseteq X$  and

$$\sum_{x \in C} \beta(x) \not\equiv a \pmod{2}. \quad (4)$$

For a partial assignment  $\beta_0: X_0 \rightarrow \{0, 1\}$  with  $|X_0| \leq k$  the  $r$ -round  $k$ -pebble game  $\mathcal{G}_k^r(V, \mathcal{C}, \beta_0)$  is played as follows:

- The game has two players called Verifier and Falsifier.
- The game is played in rounds with initial position  $\beta_0$ .
- Suppose  $\beta: X \rightarrow \{0, 1\}$  is the current position. Then the next round consists of the following steps:
  - Falsifier chooses  $x \in V \setminus X$  and  $X' \subseteq X$  such that  $|X' \cup \{x\}| \leq k$ .
  - Verifier chooses  $b \in \{0, 1\}$ .
  - The game moves to position  $\beta': X' \cup \{x\} \rightarrow \{0, 1\}$  with  $\beta'(x') = \beta(x')$  for  $x' \in X'$  and  $\beta'(x) = b$ .
- Falsifier wins a play if within the first  $r$  rounds an assignment  $\beta$  violates some XOR-constraint  $(C, a) \in \mathcal{C}$  (if  $r = 0$ , then Falsifier wins if the initial assignment  $\beta_0$  violates some constraint in  $\mathcal{C}$ ).
- Verifier wins a play if Falsifier does not win within the first  $r$  rounds.

We say Falsifier (respectively Verifier) wins the game  $\mathcal{G}_k^r(V, \mathcal{C}, \beta_0)$  if Falsifier (respectively Verifier) has a winning strategy for the game. The  $k$ -pebble game  $\mathcal{G}_k(V, \mathcal{C}, \beta_0)$  is played in the same way, but without any restriction on the number of rounds played.

The following lemma relates the pebble game  $\mathcal{G}_k^r(V, \mathcal{C}, \emptyset)$  to bounded-variable fragments of first-order logic and thereby, using Corollary 2.2, also to the Weisfeiler-Leman algorithm. (Here, we use  $\emptyset$  to denote the empty assignment, i.e., the domain  $X_0$  of the initial partial assignment  $\beta_0$  is empty.)

**Lemma 5.1** ([3, Lemma 2.1]). *Let  $k, r \in \mathbb{N}$  such that  $r > 0$  and  $k \geq 3$ . Let  $\mathcal{C}$  be a set of XOR-constraints over a universe  $V$  of arity at most  $k$ . Then the following statements are equivalent:*

- (i) *Falsifier wins the  $r$ -round  $k$ -pebble game  $\mathcal{G}_k^r(V, \mathcal{C}, \emptyset)$ .*
- (ii) *There exists a sentence  $\varphi \in \mathsf{L}_k^{(r)}$  such that  $\varphi \models \mathfrak{A}(\mathcal{C})$  and  $\varphi \not\models \mathfrak{B}(\mathcal{C})$ .*
- (iii) *There exists a sentence  $\varphi \in \mathsf{C}_k^{(r)}$  such that  $\varphi \models \mathfrak{A}(\mathcal{C})$  and  $\varphi \not\models \mathfrak{B}(\mathcal{C})$ .*

To obtain a set of XOR-constraints on which Falsifier requires a large number of rounds to win the pebble game, we proceed in two steps. First, for every sufficiently large  $\ell_{\text{hi}} \geq \ell_{\text{lo}}$ , we construct a set of XOR-constraints such that Falsifier wins the  $\ell_{\text{lo}}$ -pebble game, but still requires a linear number of rounds to win the  $\ell_{\text{hi}}$ -pebble game. This is formalized by the next lemma which forms the main technical contribution of this section.

**Lemma 5.2.** *There are absolute constants  $\ell_{\text{lo}} \geq 2$  and  $\delta > 1$  such that for every  $\ell_{\text{hi}} \geq \ell_{\text{lo}}$  and every  $r \geq 1$  there is a set of XOR-constraints  $\mathcal{C}$  of arity at most  $\ell_{\text{lo}}$  over a set  $V$  of size  $|V| \leq \delta \cdot \ell_{\text{hi}}^2 \cdot r$  such that Falsifier*

- (a) *wins the  $\ell_{\text{lo}}$ -pebble game  $\mathcal{G}_{\ell_{\text{lo}}}(V, \mathcal{C}, \emptyset)$ , but*
- (b) *does not win the  $r$ -round  $\ell_{\text{hi}}$ -pebble game  $\mathcal{G}_{\ell_{\text{hi}}}^r(V, \mathcal{C}, \emptyset)$ .*

We remark that a similar result has also been obtained in [3], but with weaker guarantees on the number of rounds required to win the  $\ell_{\text{hi}}$ -pebble game. It is exactly this improvement that allows us to obtain stronger lower bounds on the iteration number of  $k$ -WL in comparison to [3].

*Remark 5.3.* When applying Lemma 5.2, we also require that  $|V| \geq r$  which is not guaranteed by the lemma. However, if  $|V| < r$  we can simply add dummy variables that do not appear in any constraint to increase the number of variables. It is easy to see that all properties guaranteed by the lemma remain valid. In particular, the dummy variables do not affect the winning strategy for either player (if Falsifier asks for an assignment of a dummy variable, Verifier simply chooses any value; since dummy variables do not appear in any constraints this is always safe).

Afterwards, we rely on the following hardness compression lemma that reduces the number of variables while essentially maintaining the number of rounds that Falsifier requires to win the game.

**Lemma 5.4** (Berkholz, Nordström [3, Lemma 3.3]). *There is an absolute constant  $\Delta_0 \geq 1$  such that the following holds. Suppose  $\mathcal{C}$  is a set of XOR-constraints of arity at most  $p$  over a set  $V$  of size  $|V| = m$ . Also assume there are parameters  $\ell_{\text{lo}} > 0$ ,  $\ell_{\text{hi}} \geq \Delta_0 \ell_{\text{lo}}$  and  $r > 0$  such that Falsifier*

- (a) *wins the  $\ell_{\text{lo}}$ -pebble game  $\mathcal{G}_{\ell_{\text{lo}}}(V, \mathcal{C}, \emptyset)$ , but*
- (b) *does not win the  $r$ -round  $\ell_{\text{hi}}$ -pebble game  $\mathcal{G}_{\ell_{\text{hi}}}^r(V, \mathcal{C}, \emptyset)$ .*

*Let  $\Delta$  be an integer such that  $\Delta_0 \leq \Delta \leq \ell_{\text{hi}}/\ell_{\text{lo}}$  and  $(2\ell_{\text{hi}}\Delta)^{2\Delta} \leq m$ . Then there is a set of XOR-constraints  $\mathcal{D}$  of arity at most  $\Delta p$  over a set  $W$  of size  $|W| = \lceil m^{3/\Delta} \rceil$  such that Falsifier*

- (A) *wins the  $(\Delta\ell_{\text{lo}})$ -pebble game  $\mathcal{G}_{\Delta\ell_{\text{lo}}}(W, \mathcal{D}, \emptyset)$ , but*
- (B) *does not win the  $\frac{r}{2\ell_{\text{hi}}}$ -round  $\ell_{\text{hi}}$ -pebble game  $\mathcal{G}_{\ell_{\text{hi}}}^{r/(2\ell_{\text{hi}})}(W, \mathcal{D}, \emptyset)$ .*

Combining Lemmas 5.2 and 5.4, we obtain the following corollary.

**Corollary 5.5.** *There are absolute constants  $k_0 \in \mathbb{N}$  and  $\alpha, \varepsilon > 0$  such that for every  $d \geq k \geq k_0$  and every  $n \geq \alpha \cdot d^8 \cdot k^6$  there is a set of XOR-constraints  $\mathcal{C}$  of arity at most  $k$  over a set  $V$  of size  $|V| \leq n$  such that Falsifier wins the  $k$ -pebble game  $\mathcal{G}_k(V, \mathcal{C}, \emptyset)$ , but does not win the  $r$ -round  $d$ -pebble game  $\mathcal{G}_d^r(V, \mathcal{C}, \emptyset)$  for all  $r \leq n^{\varepsilon k}$ .*

*Proof.* Let  $\ell_{\text{lo}} \geq 2$  and  $\delta > 1$  denote the constants from Lemma 5.2. Also, let  $\Delta_0$  denote the constant from Lemma 5.4 and suppose without loss of generality that  $\delta, \Delta_0$  are integers and  $\Delta_0 \geq 3$ . We choose

$$k_0 := \max\{\Delta_0 \ell_{\text{lo}}, 6\ell_{\text{lo}}\}.$$

Let  $d \geq k \geq k_0$ . We set  $p := \ell_{\text{lo}}$ ,  $\ell_{\text{hi}} := d$  and  $\Delta := \lfloor \frac{k}{\ell_{\text{lo}}} \rfloor$ . We have  $\ell_{\text{hi}} = d \geq k \geq k_0 \geq \Delta_0 \ell_{\text{lo}}$  and  $\Delta_0 \leq \frac{k}{\ell_{\text{lo}}}$ . Since  $\Delta_0$  is an integer, we conclude that  $\Delta_0 \leq \Delta$ .

We define

$$n_0 := \max\{(\delta \cdot \ell_{\text{hi}}^2 \cdot (2\ell_{\text{hi}}\Delta)^{2\Delta})^{3/\Delta}, 4 \cdot \delta \cdot \ell_{\text{hi}}^3\}$$

and set  $\alpha := \max\{64 \cdot \delta \cdot \ell_{\text{lo}}^{-6}, 4 \cdot \delta\}$ . Let  $n \geq \alpha \cdot d^8 \cdot k^6$ . Using  $\Delta \geq \Delta_0 \geq 3$ , we get that

$$(\delta \cdot \ell_{\text{hi}}^2 \cdot (2\ell_{\text{hi}}\Delta)^{2\Delta})^{3/\Delta} \leq \delta \cdot \ell_{\text{hi}}^2 \cdot (2\ell_{\text{hi}}\Delta)^6 \leq 64 \cdot \delta \cdot \ell_{\text{hi}}^8 \cdot \left(\frac{k}{\ell_{\text{lo}}}\right)^6 \leq \alpha \cdot d^8 \cdot k^6$$

and  $4 \cdot \delta \cdot \ell_{\text{hi}}^3 \leq \alpha \cdot d^3$ . So in particular  $n \geq n_0$ . Let  $r$  be the maximal integer such that

$$(\delta \cdot \ell_{\text{hi}}^2 \cdot r)^{3/\Delta} \leq n. \quad (5)$$

Note that  $r \geq (2\ell_{\text{hi}}\Delta)^{2\Delta}$  since  $n \geq n_0$ . Let  $\mathcal{C}$  be the set of XOR-constraints of arity at most  $\ell_{\text{lo}}$  over a set  $V$  of size  $|V| \leq \delta \cdot \ell_{\text{hi}}^2 \cdot r$  obtained from Lemma 5.2. By adding dummy variables (see Remark 5.3), we may assume without loss of generality that  $m := |V| \geq r \geq (2\ell_{\text{hi}}\Delta)^{2\Delta}$ .

By applying Lemma 5.4, we obtain a set of XOR-constraints  $\mathcal{D}$  of arity at most  $\Delta p$  over a set  $W$  of size  $|W| = \lceil m^{3/\Delta} \rceil$  such that Falsifier

(A) wins the  $(\Delta\ell_{\text{lo}})$ -pebble game  $\mathcal{G}_{\Delta\ell_{\text{lo}}}(W, \mathcal{D}, \emptyset)$ , but

(B) does not win the  $\frac{r}{2\ell_{\text{hi}}}$ -round  $\ell_{\text{hi}}$ -pebble game  $\mathcal{G}_{\ell_{\text{hi}}}^{r/(2\ell_{\text{hi}})}(W, \mathcal{D}, \emptyset)$ .

First observe that  $\Delta p = \lfloor \frac{k}{\ell_{\text{lo}}} \rfloor \cdot \ell_{\text{lo}} \leq k$  and

$$|W| = \lceil m^{3/\Delta} \rceil \leq \lceil (\delta \cdot \ell_{\text{hi}}^2 \cdot r)^{3/\Delta} \rceil \leq n.$$

Since  $\Delta\ell_{\text{lo}} \leq k$ , it holds that Falsifier wins the  $k$ -pebble game  $\mathcal{G}_k(W, \mathcal{D}, \emptyset)$ . Moreover, Falsifier does not win the  $\frac{r}{2\ell_{\text{hi}}}$ -round  $d$ -pebble game  $\mathcal{G}_d^{r/(2\ell_{\text{hi}})}(W, \mathcal{D}, \emptyset)$ . We have that

$$(\delta \cdot \ell_{\text{hi}}^2 \cdot 2r)^{3/\Delta} \geq n$$

since  $r$  is the maximal integer to satisfy Equation (5). This implies that

$$\frac{r}{2\ell_{\text{hi}}} \geq \frac{n^{\Delta/3}}{4 \cdot \delta \cdot \ell_{\text{hi}}^3} \geq n^{\frac{1}{3} \lfloor \frac{k}{\ell_{\text{lo}}} \rfloor - 1} \geq n^{\varepsilon k}$$

for some sufficiently small absolute constant  $\varepsilon > 0$ . □

With Corollary 5.5 at hand, we are now ready to prove Theorems 1.2 and 1.5.

*Proof of Theorem 1.2.* Let  $k'_0 \in \mathbb{N}$  and  $\alpha', \varepsilon' > 0$  denote the absolute constants from Corollary 5.5.

Let  $k_0 := \max\{k'_0, 3\}$ . We set  $\alpha \geq 1$  and  $\varepsilon > 0$  in such a way that for all  $d \geq k \geq k_0$  and  $n \geq \alpha d^8 k^6$  it holds that

$$\left\lfloor \frac{n}{2} \right\rfloor \geq \alpha' (d+1)^8 (k+1)^6$$

and

$$\left(\frac{n}{2} - 1\right)^{\varepsilon' k} - d \geq n^{\varepsilon k}.$$

Now, let us fix some  $d \geq k \geq k_0$  and  $n \geq \alpha d^8 k^6$ . Let  $d' := d+1$ ,  $k' := k$  and  $n' := \lfloor \frac{n}{2} \rfloor$ . We apply Corollary 5.5 with parameters  $d', k', n'$  and obtain a set of XOR-constraints  $\mathcal{C}$  of arity at most  $k'$  over a set  $V'$  of size  $|V'| \leq n'$  such that Falsifier wins the  $k'$ -pebble game  $\mathcal{G}_{k'}(V', \mathcal{C}, \emptyset)$ , but does not win the  $r'$ -round  $d'$ -pebble game  $\mathcal{G}_{d'}^{r'}(V', \mathcal{C}, \emptyset)$  for  $r' = (n')^{\varepsilon' k'}$ .

Let  $\mathfrak{A} := \mathfrak{A}(\mathcal{C})$  and  $\mathfrak{B} := \mathfrak{B}(\mathcal{C})$ . Then  $|V(\mathfrak{A})| = |V(\mathfrak{B})| = 2|V'| \leq 2n' \leq n$ . Note that we can easily increase the size of both structures by adding isolated elements that do not participate in any relations. Also, note that both structures have arity at most  $k' = k$ .

By Lemma 5.1 and Corollary 2.2,  $k$ -WL distinguishes between  $\mathfrak{A}$  and  $\mathfrak{B}$ . On the other hand, again by Lemma 5.1 and Corollary 2.2,  $d$ -WL does not distinguish  $\mathfrak{A}$  and  $\mathfrak{B}$  after  $r := r' - d$  refinement rounds. We get that

$$r = r' - d = (n')^{\varepsilon' k'} - d \geq \left(\frac{n}{2} - 1\right)^{\varepsilon' k} - d \geq n^{\varepsilon k}. \quad \square$$

*Proof of Theorem 1.5.* This follows directly from Corollary 5.5 and Lemma 5.1.  $\square$

The remainder of this section is devoted to the proof of Lemma 5.2.

## 5.2 The Closure of the Constraint Set

The critical step in the proof of Lemma 5.2 is to argue that Verifier survives a linear number of rounds even for a large number of pebbles. Here, we rely on an alternative description of winning positions in terms of a closure operator.

Let  $k \in \mathbb{N}$ . Let  $V$  be a finite set and let  $\mathcal{C}$  be a set of XOR-constraints over  $V$  of arity at most  $k$ . We define the  $k$ -attractor

$$\text{attr}_k(\mathcal{C}) := \mathcal{C} \cup \left\{ (C_1 \oplus C_2, a_1 + a_2 \bmod 2) \mid (C_1, a_1), (C_2, a_2) \in \mathcal{C}, |C_1 \oplus C_2| \leq k \right\}.$$

Here,  $C_1 \oplus C_2$  denotes the symmetric difference between the two sets, that is,  $C_1 \oplus C_2 := (C_1 \cup C_2) \setminus (C_1 \cap C_2)$ .

Intuitively speaking, if  $x_1 + \dots + x_\ell \equiv a_1 \bmod 2$  and  $y_1 + \dots + y_m \equiv a_2 \bmod 2$  are two constraints in  $\mathcal{C}$ , then every satisfying assignment also has to satisfy the equation  $x_1 + \dots + x_\ell + y_1 + \dots + y_m \equiv a_1 + a_2 \bmod 2$ . Since all variables appearing in both sets  $\{x_1, \dots, x_\ell\}$  and  $\{y_1, \dots, y_m\}$  cancel over  $\mathbb{F}_2$ , we only need to keep those variables appearing in the symmetric difference. In the case that the resulting number of variables is bounded by  $k$ , we add the corresponding equation to the  $k$ -attractor of the constraint set.

We define  $\text{cl}_k^{(0)}(\mathcal{C}) := \mathcal{C}$  and  $\text{cl}_k^{(r+1)}(\mathcal{C}) := \text{attr}_k(\text{cl}_k^{(r)}(\mathcal{C}))$  for all  $r \geq 0$ . Finally, we define the  $k$ -closure of  $\mathcal{C}$  to be the set  $\text{cl}_k(\mathcal{C}) := \text{cl}_k^{(r)}(\mathcal{C})$  for the minimal  $r \geq 0$  such that  $\text{cl}_k^{(r+1)}(\mathcal{C}) = \text{cl}_k^{(r)}(\mathcal{C})$ .

The following lemma provides the key method to prove that Verifier can survive a certain number of rounds.

**Lemma 5.6.** *Let  $\beta: X \rightarrow \{0, 1\}$  be a partial assignment with  $|X| \leq k$  such that  $\beta$  violates no XOR-constraint  $(C, a) \in \text{cl}_k^{(r)}(\mathcal{C})$ . Then Verifier wins  $\mathcal{G}_k^r(V, \mathcal{C}, \beta)$ .*

*Proof.* We prove the statement by induction on  $r$ . For  $r = 0$  the statement is trivial. So suppose  $r \geq 1$  and Falsifier chooses  $x \in V \setminus X$  and  $X' \subseteq X$  such that  $|X' \cup \{x\}| \leq k$  in the first round. For  $b \in \{0, 1\}$  let  $\beta_b: X' \cup \{x\} \rightarrow \{0, 1\}$  be the partial assignment with  $\beta_b(x') = \beta(x')$  for  $x' \in X'$  and  $\beta_b(x) = b$ . Assume towards a contradiction that, for every  $b \in \{0, 1\}$ , there is some XOR-constraint  $(C_b, a_b) \in \text{cl}_k^{(r-1)}(\mathcal{C})$  violated by  $\beta_b$ . Observe that  $x \in C_b$  for both  $b \in \{0, 1\}$  (since otherwise  $\beta$  would violate  $(C_b, a_b)$  contradicting our assumption). Let  $C := C_0 \oplus C_1 \subseteq X$  and  $a := (a_0 + a_1) \bmod 2$ . Note that  $|C| \leq k$  since  $C \subseteq X$ . Then

$$\sum_{y \in C} \beta(y) \equiv \sum_{y \in C_0 \setminus \{x\}} \beta(y) + \sum_{y \in C_1 \setminus \{x\}} \beta(y) \equiv 1 + \sum_{y \in C_0} \beta_0(y) + \sum_{y \in C_1} \beta_1(y) \equiv 1 + a \pmod{2}$$

and  $(C, a) \in \text{cl}_k^{(r)}(\mathcal{C})$ . Hence,  $\beta$  violates some  $(C, a) \in \text{cl}_k^{(r)}(\mathcal{C})$  which is a contradiction.

So there is some  $b \in \{0, 1\}$  such that  $\beta_b$  violates no XOR-constraint in  $\text{cl}_k^{(r-1)}(\mathcal{C})$ . Verifier chooses such a  $b \in \{0, 1\}$  and the game moves to position  $\beta_b$  which violates no XOR-constraint in  $\text{cl}_k^{(r-1)}(\mathcal{C})$ . So Verifier wins  $\mathcal{G}_k^{r-1}(V, \mathcal{C}, \beta_b)$  by the induction hypothesis which implies that Verifier also wins  $\mathcal{G}_k^r(V, \mathcal{C}, \beta)$ .  $\square$

## 5.3 Layered Graphs and Expansion

Next, we discuss the construction of certain expander graphs. Overall, we are aiming to construct what we refer to as *single-neighbor layered expanders*. Towards this end, we start with constructing standard bipartite expander graphs with an expansion that is close to the minimum degree of one side of the bipartite graph. We then define *single-neighbor expanders* and observe

that bipartite expanders with large expansion also are single-neighbor expanders (with a slightly smaller expansion parameter). Finally, we obtain *single-neighbor layered expanders* by “stacking single-neighbor expanders on top of each other”.

### 5.3.1 Expander Graphs

We start by defining standard bipartite expander graphs.

**Definition 5.7.** Let  $0 < \gamma < 1$  and  $\alpha > 1$  be constants and let  $G = (V, W, E)$  be a bipartite graph. We say that  $G$  is an  $(\alpha, \gamma)$ -*expander* if for every  $\emptyset \neq Y \subseteq W$  with  $|Y| \leq \gamma|W|$  it holds that

$$N(Y) \geq \alpha|Y|.$$

For more information on expander graphs we refer to [18, 24]. The references also contain variants of the following standard argument that guarantees the existence of graphs with good expansion properties. For our purposes, the crucial property in the lemma below is that the expansion  $\alpha$  is relatively close to the degree of the vertices in  $W$ .

**Lemma 5.8.** *There is some number  $R_0 \geq 2$  such that for every  $r \geq R_0$  and every  $n \geq 4r$  there is a  $(\frac{3}{4}r, \frac{1}{20r})$ -expander  $G = (V, W, E)$  such that  $|V| = |W| = n$  and  $\deg(w) = r$  for all  $w \in W$ .*

*Proof.* Suppose  $r$  is sufficiently large. Let  $V, W$  be two sets with  $|V| = |W| \geq 4r$ . We construct a bipartite graph  $G = (V, W, E)$  using the following random process: for each  $w \in W$  we select independently and uniformly at random a set of  $r$  distinct neighbors from  $V$ . We prove that, for  $r$  sufficiently large, with positive probability the graph  $G$  is a  $(\frac{3}{4}r, \frac{1}{20r})$ -expander.

Let  $n := |V| = |W|$ . For  $X \subseteq V$  and  $Y \subseteq W$  let  $p_{X,Y}$  denote the probability that  $N(Y) \subseteq X$ . Then

$$p_{X,Y} \leq \left( \frac{|X|}{n} \right)^{r \cdot |Y|}.$$

Furthermore, let  $\alpha := \frac{3}{4}r$  and  $\gamma := \frac{1}{20r}$ . Let  $p$  be the probability that  $G$  is not a  $(\gamma, \alpha)$ -expander. Then, using the inequality  $\binom{n}{k} \leq (ne/k)^k$ , we get

$$\begin{aligned} p &\leq \sum_{\substack{Y \subseteq W \\ |Y| \leq \gamma \cdot n}} \sum_{\substack{X \subseteq V \\ |X| = \lfloor \alpha|Y| \rfloor}} p_{X,Y} \\ &\leq \sum_{s=1}^{\lfloor \gamma \cdot n \rfloor} \sum_{\substack{Y \subseteq W \\ |Y|=s}} \sum_{\substack{X \subseteq V \\ |X| = \lfloor \alpha|Y| \rfloor}} \left( \frac{|X|}{n} \right)^{r \cdot |Y|} \\ &\leq \sum_{s=1}^{\lfloor \gamma \cdot n \rfloor} \binom{n}{s} \binom{n}{\lfloor \alpha s \rfloor} \left( \frac{\alpha s}{n} \right)^{r \cdot s} \\ &\leq \sum_{s=1}^{\lfloor \gamma \cdot n \rfloor} \left( \frac{ne}{s} \right)^s \left( \frac{ne}{\alpha s} \right)^{\alpha \cdot s} \left( \frac{\alpha s}{n} \right)^{r \cdot s} \\ &= \sum_{s=1}^{\lfloor \gamma \cdot n \rfloor} \left[ \left( \frac{ne}{s} \right) \left( \frac{ne}{\alpha s} \right)^\alpha \left( \frac{\alpha s}{n} \right)^r \right]^s \\ &= \sum_{s=1}^{\lfloor \gamma \cdot n \rfloor} \left[ \left( \frac{s}{n} \right)^{r-\alpha-1} e^{1+\alpha} \alpha^{r-\alpha} \right]^s \\ &= \sum_{s=1}^{\lfloor \gamma \cdot n \rfloor} \left[ \left( \frac{s}{n} \right)^{r/4-1} e^{1+3r/4} (3r/4)^{r/4} \right]^s \end{aligned}$$

$$\leq \sum_{s=1}^{\lfloor \gamma \cdot n \rfloor} \left[ \gamma^{r/4-1} e^{1+3r/4} (3r/4)^{r/4} \right]^s.$$

Now let  $x := \gamma^{r/4-1} e^{1+3r/4} (3r/4)^{r/4}$ . For  $r$  sufficiently large we get

$$x = (20r)^{1-r/4} e^{1+3r/4} (3r/4)^{r/4} = 20er \left( \frac{3e^3}{80} \right)^{r/4} < 1/10.$$

It follows that

$$p \leq \sum_{s=1}^{\infty} x^s = \frac{x}{1-x} \leq \frac{1}{9}.$$

In particular,  $p < 1$  which implies the existence of the desired expander graph.  $\square$

Next, we turn to what we call single-neighbor expanders where each sufficiently small set  $Y \subseteq V$  is required to have a large number of neighbors that additionally satisfy the property that they are the neighbor of only a single vertex from  $Y$ . Let  $G = (V, W, E)$  be a bipartite graph. For  $Y \subseteq W$  we define

$$N^*(Y) = \{v \in N(Y) \mid |N(v) \cap Y| = 1\}.$$

**Definition 5.9.** Let  $0 < \gamma < 1$  and  $\alpha > 1$  be constants and let  $G = (V, W, E)$  be a bipartite graph. We say that  $G$  is an  $(\alpha, \gamma)$ -single-neighbor expander if for every  $\emptyset \neq Y \subseteq W$  with  $|Y| \leq \gamma|W|$  it holds that

$$|N^*(Y)| \geq \alpha|Y|.$$

We can obtain single-neighbor expanders from Lemma 5.8 by allowing some loss on the expansion parameter  $\alpha$ .

**Corollary 5.10.** *There is some number  $R_0 \geq 5$  such that for every  $r \geq R_0$  and every  $n \geq 4r$  there is a  $(\frac{1}{4}r, \frac{1}{20r})$ -single-neighbor expander  $G = (V, W, E)$  such that  $|V| = |W| = n$  and  $\deg(w) = r$  for all  $w \in W$ .*

*Proof.* Choose  $R_0 := \max(5, R'_0)$  where  $R'_0$  is the constant from Lemma 5.8 and suppose  $r \geq R_0$  and  $n \geq 4r$ . By Lemma 5.8, there is a  $(\frac{3}{4}r, \frac{1}{20r})$ -expander  $G = (V, W, E)$  such that  $|V| = |W| = n$  and  $\deg(w) = r$  for all  $w \in W$ . We claim that  $G$  is a  $(\frac{1}{4}r, \frac{1}{20r})$ -single-neighbor expander. Let  $Y \subseteq W$  with  $|Y| \leq \frac{n}{20r}$ . Then  $|N(Y)| \geq \frac{3}{4}r|Y|$ . Furthermore  $|N(Y)| = |N^*(Y)| + |\{v \in N(Y) \mid |N(v) \cap Y| \geq 2\}| \leq |N^*(Y)| + \frac{1}{2}r|Y|$  because every vertex in  $Y$  has degree  $r$ . Thus,  $|N^*(Y)| \geq \frac{1}{4}r|Y|$ .  $\square$

### 5.3.2 Layered Graphs

Now, we turn to the construction of single-neighbor layered expanders which is the main tool for constructing the desired constraint sets in the proof of Lemma 5.2. We start by defining a certain notion of layered graphs (see also Figure 1).

Let  $\ell, m \in \mathbb{N}$ . An  $(\ell \times m)$ -layered graph is a bipartite graph  $G = (V, W, E)$  for which there are partitions  $V = V_0 \uplus \dots \uplus V_\ell$  and  $W = W_1 \uplus \dots \uplus W_\ell$  such that

1.  $|V_i| = m$  for all  $i \in [0, \ell]$ ,
2.  $|W_i| = m$  for all  $i \in [\ell]$ ,
3.  $N_G(W_i) \subseteq V_{i-1} \cup V_i$  for all  $i \in [\ell]$ , and
4.  $G[V_i \cup W_i]$  is 1-regular (i.e., a matching) for all  $i \in [\ell]$ .



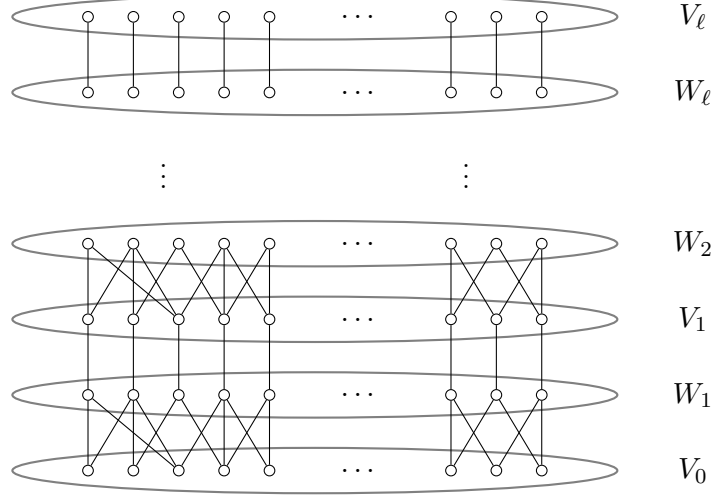


Figure 1: Visualization of  $(\ell \times m)$ -layered graphs.

With this, we are now ready to define the notion of single-neighbor layered expanders.

**Definition 5.11.** Let  $\ell, m \geq 2$ . Let  $0 < \gamma < 1$  and  $\alpha > 1$  be constants and let  $G = (V, W, E)$  be an  $(\ell \times m)$ -layered graph. We say that  $G$  is an  $(\alpha, \gamma)$ -single-neighbor  $(\ell \times m)$ -layered expander if for every  $\emptyset \neq Y \subseteq W$  with  $|Y| \leq \gamma m$  it holds that

$$N^*(Y) \geq \alpha|Y|.$$

Note that an  $(\alpha, \gamma)$ -single-neighbor  $(\ell \times m)$ -layered expander is not a  $(\alpha, \gamma)$ -single-neighbor expander since we are only considering sets  $Y \subseteq W$  of size  $|Y| \leq \gamma m$ , i.e., we are only considering sets that are smaller (by a factor of  $\gamma$ ) than a single layer of a layered graph. In particular, the reader is encouraged to think of  $\ell$  being much larger than  $m$ . In this case, such a graph is far from being a (global) expander, but the key property is that it behaves like an expander when only considering a few layers of the graph.

By again allowing some small loss on the expansion parameter  $\alpha$ , we can obtain single-neighbor layered expanders by “stacking  $\ell$  copies of a single-neighbor expander on top of each other”.

**Corollary 5.12.** *There is some number  $R_0 \geq 9$  such that for every  $r \geq R_0$ , every  $\ell \geq 1$ , and every  $m \geq 4r$  there is a  $(\frac{1}{4}r - 1, \frac{1}{20r})$ -single-neighbor  $(\ell \times m)$ -layered expander  $G = (V, W, E)$  with partitions  $V = V_0 \uplus \dots \uplus V_\ell$  and  $W = W_1 \uplus \dots \uplus W_\ell$  such that  $N_G(w) \cap V_{i-1} = r$  for all  $w \in W_i$  and all  $i \in [\ell]$ .*

*Proof.* Choose  $R_0 := \max(9, R'_0)$  where  $R'_0$  is the constant from Corollary 5.10 and suppose  $r \geq R_0$ ,  $\ell \geq 1$ , and  $m \geq 4r$ . By Corollary 5.10, there is a  $(\frac{1}{4}r, \frac{1}{20r})$ -single-neighbor expander  $G' = (V', W', E')$  such that  $|V'| = |W'| = m$  and  $\deg_{G'}(w') = r$  for all  $w' \in W'$ . Suppose  $V' = \{v'_1, \dots, v'_m\}$  and  $W' = \{w'_1, \dots, w'_m\}$ .

We set  $V_i := \{v_{i,1}, \dots, v_{i,m}\}$  for all  $i \in [0, \ell]$  and  $W_i := \{w_{i,1}, \dots, w_{i,m}\}$  for all  $i \in [\ell]$ . Also, we set

$$E := \{v_{i-1,j}w_{i,k} \mid i \in [\ell], v'_j w'_k \in E'\} \cup \{v_{i,j}w_{i,j} \mid i \in [\ell], j \in [m]\}.$$

Clearly,  $G = (V, W, E)$  is an  $(\ell \times m)$ -layered graph.

Let  $\alpha := \frac{1}{4}r$  and  $\gamma := \frac{1}{20r}$ . Also let  $Y \subseteq W$  such that  $|Y| \leq \gamma m$ . We define  $Y_i := Y \cap W_i$  for all  $i \in [\ell]$ . Observe that  $|Y_i| \leq \gamma m$  for all  $i \in [\ell]$  and  $Y_1, \dots, Y_\ell$  forms a partition of  $Y$ . Let  $\mathcal{I} := \{i \in [\ell] \mid Y_i \neq \emptyset\}$ . Since  $G'$  is an  $(\alpha, \gamma)$ -single-neighbor expander, we conclude that

$$|N^*(Y_i) \cap V_{i-1}| \geq \alpha|Y_i|$$

for all  $i \in \mathcal{I}$ . Moreover, since  $G[V_i \cup W_i]$  is 1-regular (i.e., a matching) for all  $i \in [\ell]$ , we conclude that

$$|N^*(Y) \cap V_{i-1}| \geq \alpha|Y_i| - |Y_{i-1}|$$

for all  $i \in \mathcal{I}$  (we set  $Y_0 := \emptyset$ ). So overall

$$|N^*(Y)| \geq \sum_{i \in \mathcal{I}} \alpha|Y_i| - |Y_{i-1}| \geq \sum_{i \in \mathcal{I}} (\alpha - 1)|Y_i| = (\alpha - 1)|Y|$$

as desired.  $\square$

## 5.4 Constraint Sets from Layered Expanders

Now, we turn to the construction of constraint sets from layered graphs. For a bipartite graph  $G = (V, W, E)$  we define the XOR-constraint set  $\mathcal{C}_G := \{(N(w), 0) \mid w \in W\}$  over the variable set  $V$ . Slightly abusing notation, for  $C \subseteq V$ , we shall also write  $C \in \mathcal{C}_G$  if  $(C, 0) \in \mathcal{C}_G$ .

The basic idea for the construction of the XOR-constraint set  $\mathcal{C}$  is to take a layered graph  $G = (V, W, E)$  with partitions  $V = V_0 \uplus \dots \uplus V_\ell$  and  $W = W_1 \uplus \dots \uplus W_\ell$ , and set

$$\mathcal{C} := \mathcal{C}_G \cup \{(\{x\}, 0) \mid x \in V_0\} \cup \{(\{x_\ell\}, 1)\}$$

for some arbitrary  $x_\ell \in V_\ell$ . It is not difficult to see that this constraint set is unsatisfiable. Indeed, every variable in layer  $V_0$  needs to be set to 0, and if all variables in layer  $V_{i-1}$  are set to 0, then the constraints obtained from the vertices in  $W_i$  enforce that every variable in layer  $V_i$  needs to be set to 0 as well (using that  $G[V_i \cup W_i]$  is a matching). This inductive argument can be easily turned into a winning strategy for Falsifier that requires  $O(\ell)$  many rounds (assuming the degree of all vertices in  $W$  is bounded by some absolute constant  $d \leq k$  where  $k$  denotes the number of variables available in the game).

Now, the central claim is that, if we start with a single-neighbor layered expander, this strategy is essentially optimal. Let us suppose for the moment that only constraints from  $\mathcal{C}_G$  are present and consider the  $k$ -closure  $\text{cl}_k(\mathcal{C}_G)$ . What we need to avoid is that  $\text{cl}_k(\mathcal{C}_G)$  contains some constraint that is “non-local”. For example, if  $\text{cl}_k(\mathcal{C}_G)$  would contain a constraint  $(\{x_1, x_2, x_3\}, 0)$  such that  $x_1, x_2 \in V_0$  and  $x_3 \in V_\ell$ , then Falsifier could use such a (derived) constraint to immediately conclude that certain variables in the last layer need to be set to 0 and potentially follow a different strategy to win the game faster. The main point is that, by using single-neighbor layered expanders, we ensure that all “relevant” constraints in  $\text{cl}_k(\mathcal{C}_G)$  are “local”, i.e., they can only contain variables of  $O(k)$  consecutive layers. (Here, the reader may note that if  $|N(w_1) \cup N(w_2)| \leq k$  then  $(N(w_1) \oplus N(w_2), 0)$  is always contained in the closure even if  $w_1$  and  $w_2$  are far apart. However, in such a case,  $N(w_1) \cap N(w_2) = \emptyset$  and the derived constraint  $(N(w_1) \cup N(w_2), 0)$  is not “relevant” since, whenever it is violated by a partial assignment, one of the constraints associated with  $w_1$  or  $w_2$  is also violated.) This way, even when adding all constraints from  $\text{cl}_k(\mathcal{C}_G)$  to the initial set, the best that Falsifier can do is essentially to follow the above inductive strategy (with the exception that Falsifier may skip up to  $O(k)$  layers in one step which, however, does not cause any problems for our arguments).

For technical reasons, the formal arguments slightly deviate from the intuitive ideas described above. To start, instead of working with the  $k$ -closure  $\text{cl}_k(\mathcal{C}_G)$ , it turns out to be more convenient to work with the following set.

Let  $\alpha > 1$  and  $k \geq 1$ . We define the set

$$\text{cl}_{k,\alpha}^*(\mathcal{C}_G) := \left\{ \left( \bigoplus_{D \in \mathcal{D}} D, 0 \right) \mid \mathcal{D} \subseteq \mathcal{C}_G, |\mathcal{D}| \leq \frac{k}{\alpha}, \left| \bigoplus_{D \in \mathcal{D}} D \right| \leq k \right\}.$$

We remark that, for  $\mathcal{D} = \emptyset$ , the constraint  $(\emptyset, 0)$  is added to  $\text{cl}_{k,\alpha}^*(\mathcal{C}_G)$ . Observe that  $\mathcal{C}_G \subseteq \text{cl}_{k,\alpha}^*(\mathcal{C}_G)$  if  $k \geq \alpha$  (which is always the case in our constructions). So the next lemma implies that  $\text{cl}_k(\mathcal{C}_G) \subseteq \text{cl}_{k,\alpha}^*(\mathcal{C}_G)$  if  $G$  is a suitable single-neighbor layered expander.

**Lemma 5.13.** Suppose  $\alpha > 1$  and  $0 < \gamma < 1$ . Let  $G = (V, W, E)$  be an  $(\alpha, \gamma)$ -single-neighbor  $(\ell \times m)$ -layered expander such that  $\deg(w) \leq d$  for all  $w \in W$  and suppose  $d \leq k \leq \frac{1}{2}\gamma m$ . Then

$$\text{attr}_k(\text{cl}_{k,\alpha}^*(\mathcal{C}_G)) = \text{cl}_{k,\alpha}^*(\mathcal{C}_G).$$

*Proof.* Let  $\mathcal{C}^* := \text{cl}_{k,\alpha}^*(\mathcal{C}_G)$ . Suppose  $C \in \text{attr}_k(\mathcal{C}^*)$ , that is, there are  $C_1, C_2 \in \mathcal{C}^*$  such that  $C := C_1 \oplus C_2$  and  $|C| \leq k$ . By definition, there are integers  $s, t \leq \frac{k}{\alpha}$  and  $D_1, \dots, D_s, D_{s+1}, \dots, D_{s+t} \in \mathcal{C}_G$  such that  $C_1 = D_1 \oplus \dots \oplus D_s$  and  $C_2 = D_{s+1} \oplus \dots \oplus D_{s+t}$ . Moreover,  $D_1, \dots, D_s$  are pairwise distinct as well as  $D_{s+1}, \dots, D_{s+t}$  are pairwise distinct. We have  $C = D_1 \oplus \dots \oplus D_{s+t}$ . Let

$$\mathcal{D} := \{D_1, \dots, D_s\} \oplus \{D_{s+1}, \dots, D_{s+t}\}$$

and let  $Y := \{w \in W \mid N(w) \in \mathcal{D}\}$ . Clearly,  $C = \bigoplus_{D \in \mathcal{D}} D$ . Suppose towards a contradiction that  $|\mathcal{D}| > \frac{k}{\alpha}$ . Then  $|Y| > \frac{k}{\alpha}$  and moreover,  $|Y| \leq s + t \leq 2\frac{k}{\alpha} \leq 2k \leq \gamma m$  and thus,  $|N^*(Y)| \geq \alpha|Y| > k$ . But on the other hand  $N^*(Y) \subseteq C$  which implies that  $|N^*(Y)| \leq k$ . This is a contradiction. So  $|\mathcal{D}| \leq \frac{k}{\alpha}$  which implies that  $C \in \mathcal{C}^*$  as desired.  $\square$

**Lemma 5.14.** Suppose  $\alpha > 1$  and  $0 < \gamma < 1$ . Let  $G = (V, W, E)$  be an  $(\alpha, \gamma)$ -single-neighbor  $(\ell \times m)$ -layered expander such that  $\deg(w) \leq d$  for all  $w \in W$  and suppose  $d \leq k \leq \frac{1}{2}\gamma m$ . Then  $|C| \geq 2$  for all  $C \in \text{cl}_{k,\alpha}^*(\mathcal{C}_G)$  such that  $C \neq \emptyset$ .

*Proof.* Let  $C \in \text{cl}_{k,\alpha}^*(\mathcal{C}_G)$  such that  $C \neq \emptyset$  and let  $C_1, \dots, C_s \in \mathcal{C}_G$  such that  $C = C_1 \oplus \dots \oplus C_s$  for some  $s \leq \frac{k}{\alpha} \leq k$ . Furthermore, let  $Y := \{w \in W \mid \exists i \in [s]: N(w) = C_i\}$ . Observe that  $1 \leq |Y| \leq k \leq \gamma m$ . Then  $N^*(Y) \subseteq C$  and thus,  $|C| \geq |N^*(Y)| \geq \alpha|Y| > 1$ .  $\square$

Next, we prove that Falsifier wins the pebble game if we set all variables in layer  $V_0$  to 0, and a single variable in the last layer  $V_\ell$  to 1. For technical reasons, we do not add  $(\{x_\ell\}, 1)$  to the constraint set, but rather consider an initial assignment that assigns value 1 to variable  $x_\ell$ .

**Lemma 5.15.** Let  $G = (V, W, E)$  be an  $(\ell \times m)$ -layered graph with partitions  $V = V_0 \uplus \dots \uplus V_\ell$  and  $W = W_1 \uplus \dots \uplus W_\ell$  such that  $\deg(w) \leq k$  for all  $w \in W$ . Let  $x_\ell \in V_\ell$  and suppose  $\beta_\ell: \{x_\ell\} \rightarrow \{0, 1\}$  is the partial assignment defined via  $\beta_\ell(x_\ell) = 1$ . Then Falsifier wins  $\mathcal{G}_k(W, \mathcal{C}, \beta_\ell)$  where

$$\mathcal{C} := \mathcal{C}_G \cup \{(\{x\}, 0) \mid x \in V_0\}.$$

*Proof.* We prove by induction on  $i = 0, \dots, \ell$  that Falsifier wins  $\mathcal{G}_k(V, \mathcal{C}, \beta_i)$  where  $\beta_i$  is any partial assignment for which  $\beta_i(x_i) = 1$  for some  $x_i \in V_i$ .

The base case  $i = 0$  is trivial since  $(\{x_0\}, 0) \in \mathcal{C}$  for every  $x_0 \in V_0$ . For the inductive step, suppose  $i \in [\ell]$  and consider some partial assignment  $\beta_i$  for which there is some  $x_i \in V_i$  such that  $\beta_i(x_i) = 1$ . Since  $G = (V, W, E)$  is an  $(\ell \times m)$ -layered graph, there is a unique vertex  $w_i \in W_i$  such that  $w_i x_i \in E$ . Moreover,  $N_G(w_i) \subseteq V_{i-1} \cup V_i$ . If  $N_G(w_i) = \{x_i\}$ , then  $(\{x_i\}, 0) \in \mathcal{C}_G$  and Falsifier wins immediately. So suppose that  $N_G(w_i) \cap V_{i-1} \neq \emptyset$ . Since  $\deg(w_i) \leq k$ , Falsifier can move to a partial assignment  $\beta_{i-1}: X_i \rightarrow \{0, 1\}$  where  $X_i = N_G(w_i)$  and  $\beta_{i-1}(x_i) = 1$ . If  $\beta_{i-1}$  violates the XOR-constraint  $(X_i, 0)$ , then Falsifier wins immediately. Otherwise,  $\sum_{y \in X_i} \beta_{i-1}(y) = 0$ . Together with the fact that  $\beta_{i-1}(x_i) = 1$ , this implies that there is some  $x_{i-1} \in X_i \cap V_{i-1}$  such that  $\beta_{i-1}(x_{i-1}) = 1$ . So Falsifier wins by the induction hypothesis.  $\square$

The next lemma forms the key technical lemma stating that Falsifier requires a large number of rounds to win if the constraint set is obtained from a single-neighbor layered expander.

**Lemma 5.16.** Suppose  $\alpha > 1$  and  $0 < \gamma < 1$ . Let  $G = (V, W, E)$  be an  $(\alpha, \gamma)$ -single-neighbor  $(\ell \times m)$ -layered expander with partitions  $V = V_0 \uplus \dots \uplus V_\ell$  and  $W = W_1 \uplus \dots \uplus W_\ell$  such that  $\deg(w) \leq d$  for all  $w \in W$  and suppose  $d \leq k \leq \frac{1}{2}\gamma m$ .

Let  $x_\ell \in V_\ell$  and suppose  $\beta_\ell: \{x_\ell\} \rightarrow \{0, 1\}$  is the partial assignment defined via  $\beta_\ell(x_\ell) = 1$ . Then Verifier wins  $\mathcal{G}_k^{r-1}(V, \mathcal{C}, \beta_\ell)$  where

$$\mathcal{C} := \mathcal{C}_G \cup \{(\{x\}, 0) \mid x \in V_0\}$$

and  $r := \lfloor \ell/2k \rfloor$ .

*Proof.* Let

$$\mathcal{C}_G^* := \text{cl}_{k,\alpha}^*(\mathcal{C}_G)$$

and define  $\mathcal{C}^* := \mathcal{C}_G^* \cup \{(\{x\}, 0) \mid x \in V_0\}$ . We show that Verifier wins  $\mathcal{G}_k^{r-1}(V, \mathcal{C}^*, \beta_\ell)$  which clearly implies the claim since  $\mathcal{C} \subseteq \mathcal{C}^*$  (using that  $\alpha \leq d \leq k$ ). By Lemma 5.6, it suffices to show that  $\beta_\ell$  violates no XOR-constraint from the set  $\text{cl}_k^{(r-1)}(\mathcal{C}^*)$ , or equivalently  $(\{x_\ell\}, 0) \notin \text{cl}_k^{(r-1)}(\mathcal{C}^*)$  (note that all constraints in  $\text{cl}_k^{(r-1)}(\mathcal{C}^*)$  are of the form  $(C, 0)$ ).

We define

$$\mathcal{V}_i := \bigcup_{j=0}^{2ik} V_j$$

for all  $i \in \{0, \dots, \lfloor \ell/2k \rfloor\}$ . Finally, we define

$$\mathcal{C}_i^* := \left\{ C \subseteq V \mid |C| \leq k, C = D \oplus U \text{ for some } D \in \mathcal{C}_G^*, U \subseteq \mathcal{V}_i \right\}$$

for all  $i \in \{0, \dots, \lfloor \ell/2k \rfloor\}$ .

*Claim 5.17.*  $\text{attr}_k(\mathcal{C}_i^*) \subseteq \mathcal{C}_{i+1}^*$  for all  $i \in \{0, \dots, \lfloor \ell/2k \rfloor - 1\}$ .

*Proof.* Let  $C_1, C_2 \in \mathcal{C}_i^*$  such that  $|C_1 \oplus C_2| \leq k$ . Let  $C := C_1 \oplus C_2$ . For  $j \in \{1, 2\}$  pick  $D_j \in \mathcal{C}_G^*$  and  $U_j \subseteq \mathcal{V}_i$  such that  $C_j = D_j \oplus U_j$ . Let  $U' := U_1 \oplus U_2$ . Clearly,  $U' \subseteq \mathcal{V}_i$  and  $C = D_1 \oplus D_2 \oplus U'$ .

Let  $Y_j \subseteq W$ ,  $j \in \{1, 2\}$ , be a set of vertices of size  $|Y_j| \leq \frac{k}{\alpha} < k$  such that  $D_j = \bigoplus_{w \in Y_j} N(w)$  (recall that such a set  $Y_j$  exists by the definition of  $\text{cl}_{k,\alpha}^*(\mathcal{C}_G)$ ; for  $D_j = \emptyset$  we set  $Y_j := \emptyset$ ). Then there is some  $\lambda \in \{2ik + 1, \dots, 2(i+1)k\}$  such that  $W_\lambda \cap (Y_1 \cup Y_2) = \emptyset$ . We define

$$Y_j^{<\lambda} := Y_j \cap W_{<\lambda}$$

where  $W_{<\lambda} := \bigcup_{\mu < \lambda} W_\mu$  and

$$Y_j^{>\lambda} := Y_j \cap W_{>\lambda}$$

where  $W_{>\lambda} := \bigcup_{\mu > \lambda} W_\mu$ . Moreover, let

$$C_j^{>\lambda} := \bigoplus_{w \in Y_j^{>\lambda}} N(w)$$

for both  $j \in \{1, 2\}$ . We have

$$C_j^{>\lambda} \subseteq C_j$$

because  $C_j^{>\lambda} \subseteq D_j$  (since  $W_\lambda \cap Y_j = \emptyset$ ) and  $C_j^{>\lambda} \cap U_j = \emptyset$  (since  $\lambda > 2ki$ ). Also let

$$C_{>\lambda} := C_1^{>\lambda} \oplus C_2^{>\lambda} \subseteq C.$$

Hence,  $|C_j^{>\lambda}| \leq k$  and  $|C_{>\lambda}| \leq k$ . So  $C_j^{>\lambda} \in \mathcal{C}_G^*$  for both  $j \in \{1, 2\}$ . It follows that  $C_{>\lambda} \in \mathcal{C}_G^*$  by Lemma 5.13.

Now,  $C = C_{>\lambda} \oplus U$  for some  $U \subseteq V_0 \cup \dots \cup V_{\lambda-1} \subseteq \mathcal{V}_{i+1}$ . It follows that  $C \in \mathcal{C}_{i+1}^*$ .  $\square$

Since  $\mathcal{C}^* \subseteq \mathcal{C}_0^*$  (this holds since  $(\emptyset, 0) \in \mathcal{C}_G^*$ ) it follows by induction that

$$\text{cl}_k^{(i)}(\mathcal{C}^*) \subseteq \mathcal{C}_i^* \tag{6}$$

for all  $i \in \{0, \dots, \lfloor \ell/2k \rfloor\}$  using Claim 5.17. So it only remains to prove the following claim.

*Claim 5.18.*  $\{x_\ell\} \notin \mathcal{C}_{r-1}^*$ .

*Proof.* Let  $C \in \mathcal{C}_{r-1}^*$  such that  $C \cap V_\ell \neq \emptyset$ . Also pick  $D \in \mathcal{C}_G^*$  and  $U \subseteq \mathcal{V}_{r-1}$  such that  $C = D \oplus U$  (which exist by the definition of  $\mathcal{C}_{r-1}^*$ ). We have that

$$U \subseteq \mathcal{V}_{r-1} = \bigcup_{i=0}^{2k(r-1)} V_i \subseteq \bigcup_{i=0}^{\ell-2k} V_i.$$

Let  $Y \subseteq W$  such that  $|Y| \leq \frac{k}{\alpha} < k$  and  $D = \bigoplus_{w \in Y} N(w)$ . Let  $\lambda \in [\ell]$  be the maximal number such that  $Y \cap W_\lambda = \emptyset$ . Note that  $\lambda > \ell - k$  since  $|Y| < k$ . Now let  $D' := \bigoplus_{w \in Y \cap W_{>\lambda}} N(w)$  where  $W_{>\lambda} := \bigcup_{\mu > \lambda} W_\mu$ . Then  $D' = C \cap (V_\lambda \cup \dots \cup V_\ell)$  and hence,  $|D'| \leq k$ . It follows that  $D' \in \text{cl}_{k,\alpha}^*(\mathcal{C}_G)$ . Also  $|D'| \geq 1$  since  $C \cap V_\ell \neq \emptyset$ . So  $|D'| \geq 2$  by Lemma 5.14 and thus,  $|C| \geq 2$ .  $\square$

Finally, we require one more technical lemma that allows us to add the XOR-constraint  $(\{x_\ell\}, 1)$  to the final constraint set.

**Lemma 5.19.** *Let  $k \geq 2$  and  $r \geq 1$ . Let  $V$  be a finite set and let  $\mathcal{C}$  be a set of XOR-constraints over  $V$ . Let  $x_0 \in V$  and define  $\beta_0: \{x_0\} \rightarrow \{0, 1\}$  via  $\beta_0(x_0) = 1$ . If Verifier wins  $\mathcal{G}_k^r(V, \mathcal{C}, \beta_0)$ , then Verifier also wins  $\mathcal{G}_{k-1}^r(V, \mathcal{C} \cup \{(\{x_0\}, 1)\}, \emptyset)$ .*

*Proof.* Consider a position  $\beta: X \rightarrow \{0, 1\}$  of the game  $\mathcal{G}_{k-1}^r(V, \mathcal{C} \cup \{(\{x_0\}, 1)\}, \emptyset)$ . Throughout the game, by following a winning strategy for  $\mathcal{G}_k^r(V, \mathcal{C}, \beta_0)$ , Verifier can maintain the following properties after every round  $\ell \in [0, r]$ :

- (i) If  $x_0 \in X$ , then  $\beta(x_0) = 1$ , and
- (ii) Verifier wins the game  $\mathcal{G}_k^{r-\ell}(V, \mathcal{C}, \beta')$  where  $\beta': X \cup \{x_0\} \rightarrow \{0, 1\}$  is defined via  $\beta'(x) := \beta(x)$  for all  $x \in X$  and  $\beta(x_0) := 1$ .

Observe that the condition is satisfied initially since Verifier wins  $\mathcal{G}_k^r(W, \mathcal{C}, \beta_0)$ . All positions reached this way clearly satisfy all XOR-constraints in  $\mathcal{C} \cup \{(\{x_0\}, 1)\}$  which implies that Verifier wins  $\mathcal{G}_{k-1}^r(W, \mathcal{C} \cup \{(\{x_0\}, 1)\}, \emptyset)$ .  $\square$

With this, we are ready to prove Lemma 5.2.

*Proof of Lemma 5.2.* Let  $R_0 \geq 9$  denote the constant from Corollary 5.12 and define  $\ell_{\text{lo}} := R_0 + 1$ . Let  $d := R_0$ ,  $\alpha := \frac{1}{4}d - 1 > 1$  and  $\gamma := \frac{1}{20d}$ . Let  $\ell_{\text{hi}} \geq \ell_{\text{lo}}$  and  $r \geq 1$  be given. We define  $k := \ell_{\text{hi}} + 1$ . Also, let  $m := 2 \cdot \frac{k}{\gamma} = 40dk \geq 4d$  and  $\ell := 2k(r + 1)$ .

By Corollary 5.12, there is an  $(\alpha, \gamma)$ -single-neighbor  $(\ell \times m)$ -layered expander  $G = (V, W, E)$  such that  $\deg(w) = d + 1$  for all  $w \in W$ . Let  $V_0, \dots, V_\ell$  and  $W_1, \dots, W_\ell$  denote the layers of  $G$ . Also fix some arbitrary element  $x_\ell \in V_\ell$ . We define

$$\mathcal{C} := \mathcal{C}_G \cup \{(\{x\}, 0) \mid x \in V_0\} \cup \{(\{x_\ell\}, 1)\}.$$

Note that  $\mathcal{C}$  is a set of XOR-constraints over  $V$  of arity at most  $d + 1 = \ell_{\text{lo}}$ .

To complete the proof, we show that  $\mathcal{C}$  has the desired properties. First,

$$|V| = (\ell + 1)m = (2k(r + 1) + 1)40dk \leq 8kr \cdot 40dk = 320R_0(\ell_{\text{hi}} + 1)^2r \leq \delta \cdot \ell_{\text{hi}}^2 \cdot r$$

for some suitable absolute constant  $\delta$ . Moreover, Falsifier wins the  $\ell_{\text{lo}}$ -pebble game  $\mathcal{G}_{\ell_{\text{lo}}}(V, \mathcal{C}, \emptyset)$  by Lemma 5.15. Finally, by Lemma 5.16, Verifier wins  $\mathcal{G}_k^r(V, \mathcal{C} \setminus \{(\{x_\ell\}, 1)\}, \beta_\ell)$  where  $\beta_\ell: \{x_\ell\} \rightarrow \{0, 1\}$  is the partial assignment defined via  $\beta_\ell(x_\ell) = 1$ . So Verifier wins the  $r$ -round  $\ell_{\text{hi}}$ -pebble game  $\mathcal{G}_{\ell_{\text{hi}}}^r(V, \mathcal{C}, \emptyset)$  by Lemma 5.19.  $\square$

## 6 Trading Variable Number for Quantifier Depth

In this section, we investigate tradeoffs between the number of variables and the quantifier rank of formulas used to distinguish relational structures. More concretely, suppose  $\mathfrak{A}$  and  $\mathfrak{B}$  are two structures of size  $n$  that are distinguished by  $k$ -WL. By Corollary 2.2, there is a formula  $\varphi \in \mathcal{C}_{k+1}$  such that  $\mathfrak{A} \models \varphi$  and  $\mathfrak{B} \not\models \varphi$ . Using Theorem 1.1, we may assume that  $\varphi$  has quantifier rank at most  $O(kn^{k-1} \log n)$ . In this section, we show that there are sentences  $\psi$  that distinguish between  $\mathfrak{A}$  and  $\mathfrak{B}$  with smaller quantifier rank if we are allowed to increase the number of variables by some function in  $k$ . In other words, we can show improved bounds on the number of WL-iterations required to distinguish between  $\mathfrak{A}$  and  $\mathfrak{B}$  (compared to Theorem 1.1) by increasing the dimension of the WL-algorithm.

**Theorem 6.1** (Theorem 1.6 restated). *Let  $k \geq 2$ . Let  $\mathfrak{A}$  and  $\mathfrak{B}$  be two relational structures of arity at most  $k$  such that  $n := |V(\mathfrak{A})| = |V(\mathfrak{B})|$ . Also suppose there is a sentence  $\varphi \in \mathcal{C}_{k+1}$  such that  $\mathfrak{A} \models \varphi$  and  $\mathfrak{B} \not\models \varphi$ . Let  $d := \lceil \frac{3(k+1)}{2} \rceil$ . Then there is a sentence  $\psi \in \mathcal{C}_d^{(q)}$  of quantifier rank  $q = O(k^2 \cdot n^{\lfloor k/2 \rfloor + 1} \log n)$  such that  $\mathfrak{A} \models \psi$  and  $\mathfrak{B} \not\models \psi$ .*

Toward the proof of this theorem, let us fix some  $k \geq 2$  and suppose that  $k$  is odd, i.e.,  $k = 2\ell - 1$  for some integer  $\ell \geq 2$  (this is the crucial case). Let  $\mathfrak{A}$  be a relational structure of arity at most  $k$ . We translate  $\mathfrak{A}$  into a binary structure (i.e., a structure of arity at most two)  $\text{Bin}(\mathfrak{A})$  defined as follows. The universe of  $\text{Bin}(\mathfrak{A})$  is set to

$$V(\text{Bin}(\mathfrak{A})) := (V(\mathfrak{A}))^\ell.$$

For every atomic type  $\text{typ} \in \{\text{atp}_{\mathfrak{A}}(\mathbf{v}) \mid \mathbf{v} \in (V(\mathfrak{A}))^{2\ell}\}$  (on  $2\ell$  vertices) we introduce a binary relation symbol  $R_{\text{typ}}$  and set

$$R_{\text{typ}}^{\text{Bin}(\mathfrak{A})} := \{((v_1, \dots, v_\ell), (v_{\ell+1}, \dots, v_{2\ell})) \mid \text{atp}_{\mathfrak{A}}(v_1, \dots, v_{2\ell}) = \text{typ}\}.$$

Now, the key idea behind the proof of Theorem 6.1 is to use  $d$  variables to simulate the execution of 2-WL on the binary structure  $\text{Bin}(\mathfrak{A})$ . We can then obtain the upper bound on the quantifier rank by exploiting that 2-WL stabilizes after at most  $O(n \log n)$  rounds (see Theorem 1.1).

The next lemma translates a formula that distinguishes between  $\text{Bin}(\mathfrak{A})$  and  $\text{Bin}(\mathfrak{B})$  into a formula distinguishing  $\mathfrak{A}$  and  $\mathfrak{B}$ .

**Lemma 6.2.** *Let  $\mathfrak{A}$  and  $\mathfrak{B}$  be two relational structures of arity at most  $k$ . Suppose there is a sentence  $\varphi \in \mathcal{C}_d^{(q)}$  such that  $\text{Bin}(\mathfrak{A}) \models \varphi$  and  $\text{Bin}(\mathfrak{B}) \not\models \varphi$ . Then there is a sentence  $\tilde{\varphi} \in \mathcal{C}_{d \cdot \ell}^{(q \cdot \ell)}$  such that  $\mathfrak{A} \models \tilde{\varphi}$  and  $\mathfrak{B} \not\models \tilde{\varphi}$ .*

The proof of the lemma is a standard syntactic translation (see, e.g., [21, Chapter 1.5]) and we omit the details here.

**Lemma 6.3.** *Let  $\mathfrak{A}$  and  $\mathfrak{B}$  be two relational structures of arity at most  $k$  such that  $\text{Bin}(\mathfrak{A}) \simeq_2 \text{Bin}(\mathfrak{B})$ . Then  $\mathfrak{A} \simeq_k \mathfrak{B}$ .*

*Proof.* Consider an arbitrary structure  $\mathfrak{C}$  and define  $\chi_2 := \chi_2^{(\infty)}[\text{Bin}(\mathfrak{C})]$  to be the coloring computed by 2-WL on the structure  $\text{Bin}(\mathfrak{C})$ . We define a coloring  $\chi: (V(\mathfrak{C}))^k \rightarrow C$  by setting

$$\chi(v_1, \dots, v_k) := \chi_2((v_1, \dots, v_\ell), (v_{\ell+1}, \dots, v_{k-\ell+1}, v_k, v_k)).$$

*Claim 6.4.* Suppose  $\text{atp}_{\mathfrak{C}}(v_1, \dots, v_k) \neq \text{atp}_{\mathfrak{C}}(v'_1, \dots, v'_k)$ . Then  $\chi(v_1, \dots, v_k) \neq \chi(v'_1, \dots, v'_k)$ .



*Proof.* Let  $\mathbf{typ} := \text{atp}_{\mathfrak{C}}(v_1, \dots, v_{k-1}, v_k, v_k)$ . Then  $((v_1, \dots, v_\ell), (v_{\ell+1}, \dots, v_k, v_k)) \in R_{\mathbf{typ}}^{\text{Bin}(\mathfrak{C})}$ , but on the other hand  $((v'_1, \dots, v'_\ell), (v'_{\ell+1}, \dots, v'_k, v'_k)) \notin R_{\mathbf{typ}}^{\text{Bin}(\mathfrak{C})}$ . So

$$\text{atp}_{\text{Bin}(\mathfrak{C})}((v_1, \dots, v_\ell), (v_{\ell+1}, \dots, v_k, v_k)) \neq \text{atp}_{\text{Bin}(\mathfrak{C})}((v'_1, \dots, v'_\ell), (v'_{\ell+1}, \dots, v'_k, v'_k))$$

which implies that

$$\chi_2((v_1, \dots, v_\ell), (v_{\ell+1}, \dots, v_k, v_k)) \neq \chi_2((v'_1, \dots, v'_\ell), (v'_{\ell+1}, \dots, v'_k, v'_k)).$$

This directly implies the claim.  $\lrcorner$

*Claim 6.5.*  $\chi$  is  $k$ -stable.

*Proof.* Let  $\mathbf{v}, \mathbf{v}' \in (V(\mathfrak{C}))^k$  such that  $\chi(\mathbf{v}) = \chi(\mathbf{v}')$ . Suppose  $\mathbf{v} = (v_1, \dots, v_k)$  and  $\mathbf{v}' = (v'_1, \dots, v'_k)$ . Let us write  $\mathbf{v}_1 := (v_1, \dots, v_\ell)$  for the “first half” of  $\mathbf{v}$ , and  $\mathbf{v}_2 := (v_{\ell+1}, \dots, v_k)$  for the “second half”. Note that  $\mathbf{v}_2$  has only  $\ell - 1$  entries since  $k = 2\ell - 1$ . Similarly, we define  $\mathbf{v}'_1 := (v'_1, \dots, v'_\ell)$  and  $\mathbf{v}'_2 := (v'_{\ell+1}, \dots, v'_k)$ . For  $w \in V(\mathfrak{C})$  we write  $\mathbf{v}_2 \circ w$  for the tuple  $(v_{\ell+1}, \dots, v_k, w)$  obtained from  $\mathbf{v}_2$  by appending  $w$ . The tuple  $\mathbf{v}'_2 \circ w$  is defined analogously.

Since  $\chi_2$  is 2-stable and  $\chi_2(\mathbf{v}_1, \mathbf{v}_2 \circ v_k) = \chi_2(\mathbf{v}'_1, \mathbf{v}'_2 \circ v'_k)$ , we conclude that

$$\left\{ \left\{ (\chi_2(\mathbf{v}_1, \mathbf{w}), \chi_2(\mathbf{w}, \mathbf{v}_2 \circ v_k)) \mid \mathbf{w} \in (V(\mathfrak{C}))^\ell \right\} \right\} = \left\{ \left\{ (\chi_2(\mathbf{v}'_1, \mathbf{w}), \chi_2(\mathbf{w}, \mathbf{v}'_2 \circ v'_k)) \mid \mathbf{w} \in (V(\mathfrak{C}))^\ell \right\} \right\}.$$

Using that  $\chi_2$  refines the coloring by atomic types, it follows that

$$\begin{aligned} & \left\{ \left\{ (\chi_2(\mathbf{v}_1, \mathbf{v}_2 \circ w), \chi_2(\mathbf{v}_2 \circ w, \mathbf{v}_2 \circ v_k)) \mid w \in V(\mathfrak{C}) \right\} \right\} \\ &= \left\{ \left\{ (\chi_2(\mathbf{v}'_1, \mathbf{v}'_2 \circ w), \chi_2(\mathbf{v}'_2 \circ w, \mathbf{v}'_2 \circ v'_k)) \mid w \in V(\mathfrak{C}) \right\} \right\}. \end{aligned}$$

In particular, we get that

$$\left\{ \left\{ \chi_2(\mathbf{v}_1, \mathbf{v}_2 \circ w) \mid w \in V(\mathfrak{C}) \right\} \right\} = \left\{ \left\{ \chi_2(\mathbf{v}'_1, \mathbf{v}'_2 \circ w) \mid w \in V(\mathfrak{C}) \right\} \right\}.$$

Now let  $w, w' \in V(\mathfrak{C})$  such that  $\chi_2(\mathbf{v}_1, \mathbf{v}_2 \circ w) = \chi_2(\mathbf{v}'_1, \mathbf{v}'_2 \circ w')$ . Then

$$\chi(\mathbf{v}[w/i]) = \chi(\mathbf{v}'[w'/i])$$

for all  $i \in [k]$  using again that  $\chi_2$  is 2-stable and refines the coloring by atomic types. It follows that

$$\left\{ \left\{ (\chi(\mathbf{v}[w/1]), \dots, \chi(\mathbf{v}[w/k])) \mid w \in V(\mathfrak{C}) \right\} \right\} = \left\{ \left\{ (\chi(\mathbf{v}'[w/1]), \dots, \chi(\mathbf{v}'[w/k])) \mid w \in V(\mathfrak{C}) \right\} \right\}.$$

Overall, this implies that  $\chi$  is  $k$ -stable.  $\lrcorner$

Combining both claims, we obtain that  $\chi \preceq \chi_k^{(\infty)}[\mathfrak{C}]$ . Now, we complete the proof by setting  $\mathfrak{C}$  to the disjoint union of  $\mathfrak{A}$  and  $\mathfrak{B}$ .  $\square$

*Proof of Theorem 6.1.* First suppose that  $k$  odd, i.e.,  $k = 2\ell - 1$  for some integer  $\ell \geq 2$ . Since there is a sentence  $\varphi \in \mathbf{C}_{k+1}$  such that  $\mathfrak{A} \models \varphi$  and  $\mathfrak{B} \not\models \varphi$ , we conclude that  $\mathfrak{A} \not\equiv_k \mathfrak{B}$  using Corollary 2.2. So  $\text{Bin}(\mathfrak{A}) \not\equiv_2 \text{Bin}(\mathfrak{B})$  by Lemma 6.3. By Theorem 1.1, the 2-WL algorithm distinguishes between  $\text{Bin}(\mathfrak{A})$  and  $\text{Bin}(\mathfrak{B})$  after at most  $r = O(|V(\text{Bin}(\mathfrak{A}))| \log |V(\text{Bin}(\mathfrak{A}))|) = O(\ell \cdot n^\ell \cdot \log n)$  many refinement rounds. Using Corollary 2.2 again, this means there is a sentence  $\varphi' \in \mathbf{C}_3^{(r)}$  such that  $\text{Bin}(\mathfrak{A}) \models \varphi'$  and  $\text{Bin}(\mathfrak{B}) \not\models \varphi'$ . So there is a sentence  $\psi \in \mathbf{C}_{3 \cdot \ell}^{(r \cdot \ell)}$  such that  $\mathfrak{A} \models \psi$  and  $\mathfrak{B} \not\models \psi$  using Lemma 6.2. Note that  $3\ell = 3 \cdot \frac{k+1}{2} = d$  and  $r \cdot \ell = O(\ell^2 \cdot n^\ell \cdot \log n) = O(k^2 \cdot n^{(k+1)/2} \log n)$ .

For  $k$  being even, the statement the of theorem follows by applying the first case to  $k' = k + 1$ .  $\square$

## 7 Conclusion

We obtained new upper and lower bounds for the iteration number of the WL algorithm. First, we showed that  $k$ -WL always stabilizes after at most  $O(kn^{k-1} \log n)$  rounds for all  $k \geq 2$ , which is the first non-trivial upper bound on the iteration number for  $k \geq 3$ . We complemented this result by a lower bound of  $n^{\Omega(k)}$  which improves over the previously known lower bound of  $n^{\Omega(k/\log k)}$  [3]. Finally, we also investigated tradeoffs between the dimension and the iteration number of WL. Using known characterizations of WL, our results also imply upper and lower bounds on the quantifier rank of formulas in  $C_k$  required to distinguish between two structures.

Still, several questions remain open. The first question concerns the iteration number of  $k$ -WL on graphs. The structures on which our lower bounds hold are  $n$ -element structures of arity  $\Theta(k)$  and size  $n^{\Theta(k)}$ , and the increase in arity is inherent in the hardness condensation from [3]. The best known lower bound on the iteration number of  $k$ -WL on graphs is  $\Omega(n)$  due to Fürer [5]. As an intermediate question, one can also ask for improved lower bounds in the size of the structure (i.e., the sum of the sizes of all relations), i.e., are there structures on which the iteration number of  $k$ -WL exceeds  $\Omega(m)$  where  $m$  denotes the size of the structure?

Our next question concerns the quantifier rank of formulas in  $L_k$ . While our lower bounds extend to the logic  $L_k$  (see Theorem 1.5), this is not the case for the upper bounds that crucially rely on the availability of counting quantifiers. A non-trivial upper bound of  $O(n^2/\log n)$  on the quantifier rank of formulas in  $L_3$  has been obtained in [14]. Can we also obtain improved upper bounds on the quantifier rank of formulas in  $L_k$  for  $k \geq 4$ ?

Finally, we ask for further results on tradeoffs between the variable number and the quantifier rank. Specifically, is there an integer  $d \geq 3$  such that, for all structures  $\mathfrak{A}$  and  $\mathfrak{B}$  of size  $n$  distinguished by 3-WL,  $d$ -WL distinguishes between  $\mathfrak{A}$  and  $\mathfrak{B}$  in at most  $\tilde{O}(n)$  rounds (where  $\tilde{O}(\cdot)$  hides polylogarithmic factors)? We remark that even  $d = 3$  may be a valid choice, but any  $d \geq 3$  is sufficient to obtain further tradeoffs in the spirit of Theorem 1.6.

## References

- [1] László Babai. Graph isomorphism in quasipolynomial time [extended abstract]. In Daniel Wichs and Yishay Mansour, editors, *Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2016, Cambridge, MA, USA, June 18-21, 2016*, pages 684–697. ACM, 2016. doi:10.1145/2897518.2897542.
- [2] László Babai and Péter Frankl. *Linear algebra methods in combinatorics*. University of Chicago, 2020.
- [3] Christoph Berkholz and Jakob Nordström. Near-optimal lower bounds on quantifier depth and Weisfeiler-Leman refinement steps. In Martin Grohe, Eric Koskinen, and Natarajan Shankar, editors, *Proceedings of the 31st Annual ACM/IEEE Symposium on Logic in Computer Science, LICS '16, New York, NY, USA, July 5-8, 2016*, pages 267–276. ACM, 2016. doi:10.1145/2933575.2934560.
- [4] Jin-yi Cai, Martin Fürer, and Neil Immerman. An optimal lower bound on the number of variables for graph identification. *Comb.*, 12(4):389–410, 1992. doi:10.1007/BF01305232.
- [5] Martin Fürer. Weisfeiler-Lehman refinement requires at least a linear number of iterations. In Fernando Orejas, Paul G. Spirakis, and Jan van Leeuwen, editors, *Automata, Languages and Programming, 28th International Colloquium, ICALP 2001, Crete, Greece, July 8-12, 2001, Proceedings*, volume 2076 of *Lecture Notes in Computer Science*, pages 322–333. Springer, 2001. doi:10.1007/3-540-48224-5\_27.
- [6] Martin Grohe. The quest for a logic capturing PTIME. In *Proceedings of the Twenty-Third Annual IEEE Symposium on Logic in Computer Science, LICS 2008, 24-27 June 2008, Pittsburgh, PA, USA*, pages 267–271. IEEE Computer Society, 2008. doi:10.1109/LICS.2008.11.
- [7] Martin Grohe. *Descriptive Complexity, Canonisation, and Definable Graph Structure Theory*, volume 47 of *Lecture Notes in Logic*. Cambridge University Press, 2017. doi:10.1017/9781139028868.

- [8] Martin Grohe. The logic of graph neural networks. In *36th Annual ACM/IEEE Symposium on Logic in Computer Science, LICS 2021, Rome, Italy, June 29 - July 2, 2021*, pages 1–17. IEEE, 2021. doi:10.1109/LICS52264.2021.9470677.
- [9] Martin Grohe and Sandra Kiefer. Logarithmic weisfeiler-leman identifies all planar graphs. In Nikhil Bansal, Emanuela Merelli, and James Worrell, editors, *48th International Colloquium on Automata, Languages, and Programming, ICALP 2021, July 12-16, 2021, Glasgow, Scotland (Virtual Conference)*, volume 198 of *LIPIcs*, pages 134:1–134:20. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2021. doi:10.4230/LIPIcs.ICALP.2021.134.
- [10] Martin Grohe and Oleg Verbitsky. Testing graph isomorphism in parallel by playing a game. In Michele Bugliesi, Bart Preneel, Vladimiro Sassone, and Ingo Wegener, editors, *Automata, Languages and Programming, 33rd International Colloquium, ICALP 2006, Venice, Italy, July 10-14, 2006, Proceedings, Part I*, volume 4051 of *Lecture Notes in Computer Science*, pages 3–14. Springer, 2006. doi:10.1007/11786986\_2.
- [11] Neil Immerman and Eric Lander. Describing graphs: A first-order approach to graph canonization. In Alan L. Selman, editor, *Complexity Theory Retrospective: In Honor of Juris Hartmanis on the Occasion of His Sixtieth Birthday, July 5, 1988*, pages 59–81. Springer New York, New York, NY, 1990. doi:10.1007/978-1-4612-4478-3\_5.
- [12] Sandra Kiefer and Brendan D. McKay. The iteration number of colour refinement. In Artur Czumaj, Anuj Dawar, and Emanuela Merelli, editors, *47th International Colloquium on Automata, Languages, and Programming, ICALP 2020, July 8-11, 2020, Saarbrücken, Germany (Virtual Conference)*, volume 168 of *LIPIcs*, pages 73:1–73:19. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2020. doi:10.4230/LIPIcs.ICALP.2020.73.
- [13] Sandra Kiefer and Daniel Neuen. The power of the weisfeiler-leman algorithm to decompose graphs. *SIAM J. Discret. Math.*, 36(1):252–298, 2022. doi:10.1137/20m1314987.
- [14] Sandra Kiefer and Pascal Schweitzer. Upper bounds on the quantifier depth for graph differentiation in first-order logic. *Log. Methods Comput. Sci.*, 15(2), 2019. doi:10.23638/LMCS-15(2:19)2019.
- [15] Moritz Lichter, Iliia Ponomarenko, and Pascal Schweitzer. Walk refinement, walk logic, and the iteration number of the Weisfeiler-Leman algorithm. In *34th Annual ACM/IEEE Symposium on Logic in Computer Science, LICS 2019, Vancouver, BC, Canada, June 24-27, 2019*, pages 1–13. IEEE, 2019. doi:10.1109/LICS.2019.8785694.
- [16] Christopher Morris, Yaron Lipman, Haggai Maron, Bastian Rieck, Nils M. Kriege, Martin Grohe, Matthias Fey, and Karsten M. Borgwardt. Weisfeiler and Leman go machine learning: The story so far. *CoRR*, abs/2112.09992, 2021. URL: <https://arxiv.org/abs/2112.09992>, arXiv:2112.09992.
- [17] Christopher Morris, Martin Ritzert, Matthias Fey, William L. Hamilton, Jan Eric Lenssen, Gaurav Rattan, and Martin Grohe. Weisfeiler and Leman go neural: Higher-order graph neural networks. In *The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019, Honolulu, Hawaii, USA, January 27 - February 1, 2019*, pages 4602–4609. AAAI Press, 2019. doi:10.1609/aaai.v33i01.33014602.
- [18] Rajeev Motwani and Prabhakar Raghavan. *Randomized Algorithms*. Cambridge University Press, 1995. doi:10.1017/cbo9780511814075.
- [19] Daniel Neuen. Isomorphism testing parameterized by genus and beyond. In Petra Mutzel, Rasmus Pagh, and Grzegorz Herman, editors, *29th Annual European Symposium on Algorithms, ESA 2021, September 6-8, 2021, Lisbon, Portugal (Virtual Conference)*, volume 204 of *LIPIcs*, pages 72:1–72:18. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2021. doi:10.4230/LIPIcs.ESA.2021.72.
- [20] Daniel Neuen. Isomorphism testing for graphs excluding small topological subgraphs. In Joseph (Seffi) Naor and Niv Buchbinder, editors, *Proceedings of the 2022 ACM-SIAM Symposium on Discrete Algorithms, SODA 2022, Virtual Conference / Alexandria, VA, USA, January 9 - 12, 2022*, pages 1411–1434. SIAM, 2022. doi:10.1137/1.9781611977073.59.
- [21] Martin Otto. *Bounded Variable Logics and Counting: A Study in Finite Models*, volume 9 of *Lecture Notes in Logic*. Cambridge University Press, 2017. doi:10.1017/9781316716878.

- [22] Nino Shervashidze, Pascal Schweitzer, Erik Jan van Leeuwen, Kurt Mehlhorn, and Karsten M. Borgwardt. Weisfeiler-Lehman graph kernels. *J. Mach. Learn. Res.*, 12:2539–2561, 2011. URL: <https://dl.acm.org/doi/10.5555/1953048.2078187>.
- [23] Xiaorui Sun and John Wilmes. Faster canonical forms for primitive coherent configurations: Extended abstract. In Rocco A. Servedio and Ronitt Rubinfeld, editors, *Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing, STOC 2015, Portland, OR, USA, June 14-17, 2015*, pages 693–702. ACM, 2015. doi:10.1145/2746539.2746617.
- [24] Salil P. Vadhan. Pseudorandomness. *Found. Trends Theor. Comput. Sci.*, 7(1-3):1–336, 2012. doi:10.1561/04000000010.
- [25] Oleg Verbitsky. Planar graphs: Logical complexity and parallel isomorphism tests. In Wolfgang Thomas and Pascal Weil, editors, *STACS 2007, 24th Annual Symposium on Theoretical Aspects of Computer Science, Aachen, Germany, February 22-24, 2007, Proceedings*, volume 4393 of *Lecture Notes in Computer Science*, pages 682–693. Springer, 2007. doi:10.1007/978-3-540-70918-3\\_58.
- [26] Boris Weisfeiler and Andrei Leman. The reduction of a graph to canonical form and the algebra which appears therein. *NTI, Series 2*, 1968. English translation by Grigory Ryabov available at [https://www.itl.zcu.cz/wl2018/pdf/wl\\_paper\\_translation.pdf](https://www.itl.zcu.cz/wl2018/pdf/wl_paper_translation.pdf).
- [27] Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. How powerful are graph neural networks? In *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*. OpenReview.net, 2019. URL: <https://openreview.net/forum?id=ryGs6iA5Km>.
- [28] Alexander Zimmermann. *Representation theory*, volume 19 of *Algebra and Applications*. Springer, Cham, 2014. A homological algebra point of view. doi:10.1007/978-3-319-07968-4.